Theses and Dissertations   1. Thesis and Dissertation Collection, all items

1990-03

The characteristics of user-generated passwords

Sawyer, Darren Antwon

Monterey, California: Naval Postgraduate School

http://hdl.handle.net/10945/34860

NAVAL POSTGRADUATE SCHOOL
Monterey, California

AD-A225 390

THESIS

THE CHARACTERISTICS OF
USER-GENERATED PASSWORDS

by

Darren Antwon Sawyer

March, 1990

Thesis Co-Advisor: Moshe Zviran
Thesis Co-Advisor: William J. Haga

Approved for public release; distribution is unlimited.

90 08 16 008

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School	6b. OFFICE SYMBOL (If applicable) 37	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c. ADDRESS (City, State, and ZIP Code) Monterey CA 93943-5000		7b. ADDRESS (City, State, and ZIP Code) Monterey CA 93943-5000

8a. NAME OF FUNDING / SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

THE CHARACTERISTICS OF USER-GENERATED PASSWORDS (Unclassified)

12. PERSONAL AUTHOR(S)
Sawyer, Darren A.

13a. TYPE OF REPORT Master's Thesis	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) March 1990	15. PAGE COUNT 109

16. SUPPLEMENTARY NOTATION The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Passwords, Computer Security, User-Generated Passwords, Information System Security

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Access control based on the verification of a person's identity is commonly used in information system/computer installations. The most widely used mechanism for access control to information systems is passwords. Passwords can be machine-generated using a list of words stored in a memory bank, machine-generated using a sophisticated algorithm to create a pseudo-random combination of characters or they can be user-generated. User-generated passwords typically take on the characteristics of some type of meaningful detail that is simple in structure and easy to remember.

Memorability and security pose a difficult trade-off in password generation. On one hand a system security administrator wants passwords that are unpredictable, frequently changed and provide the greatest degree of system security achievable. Users, on the other hand, want passwords that are simple and easy to remember. When they become difficult to guess, they may become difficult to remember. When they become difficult to remember they are likely to be written down. Once written down a compromise to security —

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL William Haga/Moshe Zviran	22b. TELEPHONE (Include Area Code) (408) 646-3094	22c. OFFICE SYMBOL 54

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted.
All other editions are obsolete

☆ U.S. Government Printing Office: 1986—606-24.

19. Continued

occurs because users tend to store them in insecure places.

This thesis looks at user-generated password characteristics. Of particular interest is how password selection, memorability and predictability are affected by the number of characters in a password, the importance and sensitivity of a user's data, a user's work location, how a password was chosen, the frequency of changing a password and the frequency of logging on to a system with a password. *Theses, (EDC)*

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	

By	
Distribution/	
Availability Codes	

Dist	Avail and/or Special
A-1	

THE CHARACTERISTICS OF USER GENERATED PASSWORDS

by

Darren A. Sawyer

Lieutenant, United States Navy Reserve

B.A., Virginia Military Institute 1985

Submitted in partial fulfillment

of the requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS

from the

NAVAL POSTGRADUATE SCHOOL

MARCH 1990

Author: _____
Darren A. Sawyer

Approved by: _____
Moshe Zviran, Thesis Co-Advisor

_____
William J. Haga, Thesis Co-Advisor

_____
David R. Whipple, Chairman
Department of Administrative Sciences

ABSTRACT

Access control based on the verification of a person's identity is commonly used in information system/computer installations. The most widely used mechanism for access control to information systems is passwords. Passwords can be machine-generated using a list of words stored in a memory bank, machine-generated using a sophisticated algorithm to create a pseudo-random combination of characters or they can be user-generated. User-generated passwords typically take on the characteristics of some type of meaningful detail that is simple in structure and easy to remember.

Memorability and security pose a difficult trade-off in password generation. On one hand a system security administrator wants passwords that are unpredictable, frequently changed and provide the greatest degree of system security achievable. Users, on the other hand, want passwords that are simple and easy to remember. If passwords are chosen to make them difficult to guess, they may become difficult to remember. When they become difficult to remember they are likely to be written down. Once written down a compromise to security occurs because users tend to store them in insecure places.

This thesis looks at user-generated password characteristics. Of particular interest is how password selection, memorability and predictability are affected by the number of characters in a password, the importance and sensitivity of a users data, a users work location, how a password was chosen, the frequency of changing a password and the frequency of logging on to a system with a password.

iv

TABLE OF CONTENTS

I. INTRODUCTION

A. INFORMATION SYSTEMS RESOURCES PROTECTION

Concerns of information system managers include the security, privacy and integrity of information system resources. With the advent of timesharing, networking, distributed systems and an increased level of computer literacy, these concerns have grown considerably. A dilemma in designing computer-based information systems that provide a high degree of security is "How can we make systems that are easy to use but hard to misuse" (Smith, 1987)?

Concern for data security will take different forms in different applications. Individual users of computers may be concerned with personal privacy and wish to limit access to private data files. Commercial organizations may seek to protect data related to proprietary interest. Military agencies may be responsible for safeguarding data critical to national security (Smith, 1987).

Systems without security controls are vulnerable to fraud, industrial espionage, sabotage, alteration and disclosure. Statistical evidence indicates that most unauthorized access attempts go unnoticed. One out of 100 computer crimes is detected, one out of 22,000 is prosecuted, 33 of these prosecutions leads to a conviction (Hagopian, 1987).

Absolute security is unattainable (Kochanski, 1989). However, there are layers of security that when used in concert provide a high degree of system security. Physical security, the outermost layer, is primarily concerned with preventing access to buildings, terminal rooms and computer hardware (Durr and Gibbs, 1989). The inner layers of security are concerned with logical security - the methods that cover control of access to system resources and services (Durr and Gibbs, 1989).

Access control is required at various levels. These levels include access to buildings, terminal rooms, files and databases. At each level a certain degree of user

1

identification, authentication and authorization must be veiified. Normally physical security devices are used at physical access points. For example, cards or keys are used for entry into buildings and terminal rooms. However, other access mechanisms are required to control access to files and data. Password systems are the first line of defense that can prevent, deter and detect abusive acts to files and data (Wood, 1983). They are one of the most cost-effective computer resource access control mechanisms being widely used. However, through the years passwords were found to be lacking. Passwords that yielded a high degree of system security were found to be difficult to remember. Conversely, passwords that were easy to remember were found to yield a low degree of system security (Barton and Barton, 1984).

A 1979 study by Morris and Thompson investigated the issues of password security on a UNIX-based remotely accessed time sharing system. The study revealed the compromise between extreme security and ease of use. Its findings suggested that user-generated passwords are typically easy to guess because of their character length (2-6 characters) and the method in which they were chosen (something meaningful to the user). Passwords should be both easy to remember and difficult to guess. This trade off is difficult to achieve.

Following Morris and Thompson's (1979) work, this thesis focuses on the identifying the characteristics of user-generated passwords in the personal computer era. Emphasis is placed on an extensive look into these characteristics and how they are affected by variables that may influence password selection, memorability and predictability.

II. INFORMATION SYSTEM SECURITY: AN OVERVIEW

A. INFORMATION SECURITY: DEFINITION

Information security refers to the technological safeguards and managerial procedures which can be applied to information systems hardware, software and data to assure that organizational assets and individual privacy are protected (Hoffman, 1977). Technological safeguards are implemented on hardware, software and communication mediums; managerial safeguards include procedural controls, personnel controls and management controls, all of which, when imbedded in an administrative framework, ensure information system security (Ware, 1981).

Effective information system security requires the following steps (Hoffman, 1977):

1. risk analysis;

2. identifying and implementing required technical safeguards;

3. identifying and implementing required administrative safeguards;

4. periodic review of safeguards.

Risk analysis provides the basis for identifying required technical and administrative safeguards for information system security. Since absolute security seems to be unattainable (Kochanski, 1989), risk analysis produces a degree of security commensurate with the information to be protected and with the amount of resources to be expended (Hoffman, 1977). Pfleeger (1988) identifies six basic steps of risk analysis:

1. identify assets;

2. determine vulnerabilities;

3. estimate likelihood of exploitation;

3

4. compute expected annual loss;

5. survey applicable controls and their costs;

6. project annual savings of controls.

The mix of technical and administrative safeguards are determined by the sensitivity and importance of the information/data to be protected. The higher the sensitivity and importance the greater the requirements are for effective technical and administrative safeguards. Both safeguards, when used in concert, complement each other; one without the other leads to insecure information system resources (Hoffman, 1977).

Computer based information systems consist of six categories of resources: hardware, software, communication facilities, data, information and people. Each of these resources may be the means by which a system is compromised. Periodic reviews of the safeguards are required to ensure information system security. Reviews should be conducted when new hardware or services are implemented, after a reorganization and high personnel turnover. Spot reviews should also be conducted.

Two general approaches may be used in developing a security system: all resources are protected or only critical resources are protected (Wood, 1983). Some information managers emphasize the value of computer hardware rather than the value of the information stored in the system (Wood, 1983). Information system security affects an organization as a whole. For this reason, security has become a management issue. Information is a key to any organization's success. Information stored in computer-based systems must be protected against accidental or intentional disclosure, fraud and abuse. In order for information integrity to be maintained a certain degree of security and access control must be implemented system and organization wide. However,

security creates accessibility problems. Security may threaten the survivability of an organization if the trade-off between accessibility and protection of information resources is not optimized. An information system with no accessibility is useless to an organization, however, a system storing sensitive and vital information with no security may render an organization useless if the information is accessed by the wrong people.

Most often, when security is implemented system users view it as restricting. No longer can they access various databases without authorized access; no longer can they enter controlled spaces to browse files and manipulate data. An organization is affected in four ways by the implementation of information system security (Hagopian, 1987):

1. additional job responsibilities are assigned causing possible organizational friction;
2. the security system makes sign-on more difficult;
3. access to resources are restricted;
4. the choice of which terminal to use will be reduced.

B. JUSTIFICATION FOR INFORMATION SYSTEM SECURITY

As the use of information systems spreads so does the potential for abuse of information systems. Abuse consists of intentional disclosure, fraud, espionage, embezzlement, sabotage and larceny (Parker and Nycum, 1984). Although the number of computer crimes increases as the use and complexity of computer systems grow, many crimes go unnoticed, unreported and unprosecuted. One out of eight computer abuses is detected (Hoffer and Straub, 1989). Of those detected 32% are discovered by accident, 45% by normal system controls, eight percent by computer security officers and 4.5% by auditors (Hoffer and Straub, 1989). The cost of computer crimes results in significant monetary losses to many organizations. In their study, Hoffer and Straub

5

(1989) revealed 211 out of 1,211 firms reported 259 separate incidents of information system abuse. Five firms reported more than $100,000 in losses, one reported a two million dollar loss. Equity Funding Insurance suffered a $200 million loss due to fraud; a $21 million bank embezzlement in Los Angeles, a $10 million funds transfer fraud in Los Angeles, a $53 million security fraud in Florida and a $67 million inventory fraud in New York, are record breaking cases of computer crimes (Parker and Nycum, 1974). While some view information system security as an inconvenience and costly, one can also appreciate that the cost of implementing a security system is minimal compared to the cost of doing without it (Rash,1989).

Computer crimes result from systems resources being exposed either intentionally or unintentionally to those not authorized access to them. Exposure represents possible loss or harm while vulnerability is considered the weakness that might be exploited (Pfleeger, 1989). Exposures increase as vulnerability increases; they both can occur in the forms of disclosure, modification or destruction of information system resources (Fisher, 1984). Hoffer and Straub (1989) identified five categories of exposure :

1. unauthorized use of computer service;

2. disruption of computer service;

3. abuse to data;

4. abuse to programs;

5. abuse to hardware.

Exposures and vulnerabilities exist as a result of several factors some which act alone while others act in combination to cause catastrophic abusive acts. These factors include people, hardware, software, communication, procedures and acts of God (Fisher, 1984).

6

1. People

The people threat is the major cause of exposure. The majority of the people that create the people problem are application programmers, clerks, students, managers and systems analysts (Hoffer and Straub, 1989). 90% of computer crimes is instigated by insiders (Cooper, 1989). Many of these are people with special privileges, have access to special information who fall into the categories of employees or ex-employees and by virtue of their employment have a special measure of trust (Cooper, 1989). Motivation for abusing information systems fall into four categories (Cooper, 1989; Hoffer and Straub, 1989; Ware, 1984):

1. personal gain (money) - 30%

2. ignorance of proper conduct - 27%

3. misguided playfulness - 21%

4. maliciousness - 12%.

2. Hardware

Hardware-related exposures may be caused by inadequate or incorrect microcoding causing a legitimate request to yield unauthorized information.

3. Software

The use of software to reveal information is probably the second major cause of exposures. During software development, a common practice is to implement specific ways for a developer to quickly gain access to certain segments of a program. These quick-entry mechanisms or "back doors" may not be completely eliminated before a

7

program is released to users, allowing intruders to gain access to and modify original code.

4. Communications

With the proliferation of personal computers and their ability to communicate with other computer systems anywhere in the world, the complexity of communications security is a significant problem. No longer can a security manager confidently establish boundaries around a system. With the advent of networks, distributed systems and technological advances in communications, remote access can be achieved from anywhere through the use of a personal computer, a modem and a communications program.

5. Procedures

Procedures that have been poorly thought out can have detrimental effects. Procedures are a form of an administrative safeguard employed to implement and improve information system security. However, procedures that are poorly thought out, with little management enforcement and involvement and fail to support organizational goals and missions, will be easily circumvented, ignored and abused.

6. Acts of God

Acts of God include natural disasters such as floods, hurricanes, earthquakes or fires and can result in the loss of facilities and data. Backup and recovery procedures plus establishment of a geographically separated secondary facility can alleviate these possibilities.

C. METHODS OF DEFENSE

Protection of computer-based information systems start with commitment from an organization. This commitment includes money and management enforcement of established policies and regulations. Protection of computer-based information systems may be thought of as a layered approach, resembling an onion. Layers of security surround the resource that needs to be secured. Each layer insulates the subject and makes it more difficult to access in any way other than those planned for (Durr and Gibbs, 1989). Physical security is the outer layer, concerned with preventing access to the hardware (Durr and Gibbs, 1989). The inner layers of security are concerned with logical security - methods that cover control of access to the system resources and services (Durr and Gibbs, 1989). Each layer uses a different methodology to address the problems unique to that particular layer. The synergism of multiple layers creates a security system that protects its resources and services.

Hsiao (1979) delineates five types of defenses: operational security, physical security, hardware security, cryptographic transformations and operating system security.

1. Operational Security

The broad category of operational security encompasses two major areas: operating environment and authorization control (Hsiao, 1979).

a. Operating Environment.

An operating environment is defined in terms of the degree of access allowed to a computer system. Three possibilities exist: closed, open or unlimited (Hsiao, 1979). In a closed system only a few users have access. In an open system, any person can gain access by identifying himself or herself personally to another person authorized

to grant access. In an unlimited environment, any person can gain access with little effort.

b. *Authorization Control.*

Authority to grant access to a system can be divided into three categories: centralized, hierarchial decentralized and individual (Hsiao, 1979). Under centralized control, a person or department controls who is granted authorization. In hierarchial decentralization, functional managers have the power to grant access for specific areas under their control. The complete decentralization of control results in individual control: an owner of information is responsible to control access to it. Authorization control or authentication can take many forms from passwords to the confirmation of biological traits. Various types of authorization control are discussed later in this paper.

2. Physical Security

Physical security encompasses acts of God, man-made disasters and intrusion (Hsiao, 1979). Acts of God, such as fires and floods, may be controlled by installation of sensors and automatic suppressant systems such as a HALON 1211 fire fighting system. Man-made disasters or equipment failures such as a disk head crash can be minimized through a backup and recovery system. Intrusion, either intentional or unintentional, is a primary concern of physical security. Prior to the proliferation of network communications systems, avoiding intrusion meant keeping a person from physically entering a computer facility. With networks and distributed systems and the current ability to access a computer through remote terminals, physical security must now be concerned with preventing access through communications media. Cipher locks, identification cards and door monitors are examples of tools used for physical security.

10

3. Hardware Security

Closely related to the design of hardware is the design of the hardware security system. Various hardware components require protection from both the user and computer applications or processes desiring to use the hardware resources. Examples of tools used are special microchips called registers and operating system software.

4. Cryptographic Transformations

A different approach to security is the encoding of user access information. The underlying assumption is that intruders will be able to gain access. Rather than try to prevent access, emphasis is placed on encrypting or scrambling the data making it unusable by outsiders (Hsiao, 1979). Data that can not be interpreted are of little value. Tools commonly used are encryption and decryption algorithms.

5. Operating System Security

An operating system is the master program that controls the execution of all other processes and stays resident in main memory. Prior to the running of an application program, an operating system must be executed. An operating system acts as the mediator between competing processes and allocates resources based on demands. Gaining access to an operating system can lead to access of other programs.

D. DEFENSE TOOLS

Two of the major defense tools used in computer security are authentication and encryption (Wood, 1983).

1. Authentication

The most widely used defense tool is use of authentication methods. Identification by authentication is approached in two ways: use of natural properties, such as fingerprints, or use of artificial measures, such as passwords or magnetic cards (Ahituv et al., 1987). Authentication methods use something known (a password), something possessed (a personal key), something to be performed (a signature) or some biological trait (a fingerprint) (Fisher, 1984).

The underlying logic of authentication devices takes two forms: make computers more like people by equipping them with biometric readers or make people more like computers by equipping them with personal computerized authentication devices (Spender, 1987). Authentication devices take the form of biometrics, directly connected token reading devices (keyholes which accept electronic keys), user interface tokens (pocket devices that can generate one-time passwords) and fixed password devices (plastic cards that contain access codes read electronically) (Spender, 1987).

2. Encryption

Encryption is the rendering of information unintelligible by effecting a series of transformations through the use of variable elements (Wood, 1977). It is the process of encoding a message so that the meaning of the message is not obvious (Pfleeger, 1988). Encryption can be applied to databases, files, system applications and passwords used as authentication mechanisms. Encryption of passwords can be accomplished by

three methods: encrypting a password table stored in memory, using one-way encrypted passwords and using a personal key device that contains an encrypted code after the plain text password has been entered (Ahituv et al., 1987). The underlying objective of is to make encryption easy and decryption difficult (Pfleeger, 1988). Encryption raises the time and effort required to break a users encrypted password (Menkus, 1988).

Encryption and authentication are highly viable tools for defending against potential intruders. Authentication can be achieved through the use of passwords. Passwords are a cost-effective authentication and access control mechanism. With this background in information system security, Chapter III explores the use of passwords as a security mechanism.

III. PASSWORDS AS A SECURITY MECHANISM

A. BACKGROUND

One of the most important aspects in a secure computer system is access control. At the heart of access control is user identification (Avarne, 1988). The first step to establishing the identity of a user is through authentication. Passwords are an identity authenticator (Avarne, 1988; Porter, 1982).

1. Authentication vs Authorization

Verifying the identity of a user is a critical factor in establishing the first line of defense of an information system. Authentication verifies that a person is who he or she claims to be (Hoffman, 1977; Menkus, 1988; Porter, 1982). Authentication falls into three categories (Cooper, 1989; Menkus, 1988; Porter, 1982; Wood, 1977):

1. something possessed, e.g., card, key, special terminal;

2. something characteristic of person, e.g., biometrics, hand geometry, signature;

3. something known, e.g., password algorithm.

Clearly, authentication must take place to ensure proper identification of a system user and to facilitate the enforcing of a system security.

Once authentication has been conducted, authorization of a given user to access specific resources can be carried out. Authorization is defined as determining whether a person or object is legitimately entitled to a protected resource (Hoffman, 1977). Both authentication and authorization are used together to ensure proper access

14

control. Passwords, the most common authentication mechanism, provide a means for controlling access to system resources.

2. Definition of Passwords

The use of passwords is one the oldest method of access control. Passwords consist of a sequence of numbers, special symbols or control characters used to authenticate a user's identity (Wood, 1983). They are mutually agreed upon code words assumed to be known only by a user, a system and a system administrator. The risk of granting access to an invalid user must be measured against the cost of designing, implementing and maintaining an adequate security system. The advent of networks, distributed systems and end-user computing brought computer resources out from under the centrally protected facilities. A rudimentary password architecture became the answer to the problem of protecting dispersed resources and controlling access to these resources. Passwords offer the benefits of being relatively inexpensive, readily implementable and supported by most operating systems (Spender, 1987). A fourth benefit of adopting a password security system is familiarity. Passwords are a known methodology (Wood, 1977).

3. Objectives of Password Controls

The objective of a password is to authenticate a user of computer resources (Wood, 1983). The underlying goal is to provide a password security system at minimal inconvenience to users of a system (Morris and Thompson, 1979). As the first line of defense, password controls should prevent unauthorized users from gaining access to a system as well as preventing authorized users from engaging in activities for which they have not received permission (Morris and Thompson, 1979; Wood, 1983). Simply put, password systems should be designed to prevent, detect and deter (Wood, 1983).

15

Password systems should be flexible, user-friendly, reliable, effective and secure (Cooper, 1989). However, a trade-off exists between user-friendliness and security. The user of a password system wants an easy to remember password, an easy to use system and no restrictions to resources. An organization wants sensitive and vital information protected against espionage, modification and destruction. Protection of personal privacy, proprietary interest, administrative confidentiality (Barton and Barton, 1984) and, in the military, safeguarding data critical to national security (Smith, 1987) can be achieved through passwords. The trade-off between ease of use (user-friendliness and flexibility) and security (reliability and effectiveness) is directly applicable to passwords. Too much emphasis placed on ease of use may hamper the degree of security provided, while too much emphasis on security and controlling access may stifle organizational growth from lack of access to sufficient information in a timely manner and create resentment to toward use of the system. Although an optimum trade-off between users' interest and organizational protection of sensitive resources is difficult to achieve, the critical objective of a password security scheme is to establish that the system is interacting with the person who is purported to be on the system and access to only authorized resources is permitted (Wood, 1983).

B. TYPES OF PASSWORDS

Passwords are categorized by two methods: generation and use. Generation methods include system, user and manufacturer. Use methods include primary, secondary and duress.

1. System-Generated Passwords

System-generation of passwords is managed by a system security administrator (Menkus, 1988). An administrator's responsibilities include selection of new passwords, distribution of passwords, monitoring to ensure proper use of passwords and disposition of expired passwords. System-generated passwords are normally generated either through a random number generator or a nonsense string generator (Menkus, 1988).

The advantage of system-generated passwords is that a user is removed from the selection process. User-generated passwords are normally connected with a user's lifestyle. That makes them vulnerable to guessing by outsiders (Menkus, 1988). System-generated passwords will normally contain random characters and are not related to a user's lifestyle.

Disadvantages of system-generated passwords include difficulty in remembering, possible repetition of generation cycles, vulnerability of storage tables and the removal of a user from the selection process. Nonsensical strings of characters make guessing difficult, but also make remembrance by a user difficult. Complicated passwords tend to be forgotten or written down (Ahituv et al., 1987).

To combat this problem, some systems generate character strings that include vowels, making the strings more pronounceable and therefore memorable. The tradeoff in making system generated passwords pronounceable is that the passwords are more vulnerable (Kurzban, 1983).

2. User-Generated Passwords

User-generated passwords tend to be simple. They commonly take on the characteristics of some type of meaningful detail to the user. Typically, user-generated passwords are composed of birth dates, spouse's names, nicknames, street names and other data connected with a user's lifestyle (Menkus, 1988). In many cases, passwords can be found in personnel files. The Department of Defense uses teams of computer experts to test the integrity of security systems. These tiger teams routinely comb personnel files, for passwords based on personal data. They are usually successful (Wood, 1983).

Morris and Thompson (1979) studied the characteristics of user-generated passwords. They collected 3,289 users' passwords over a long period. They wanted to identify a user's habits in selecting password when no constraints were placed on their choice. The distribution of these passwords based on their physical characteristics is shown below:

1. 0.45% were a single ASCII character;

2. 2.0% were strings of two ASCII characters;

3. 14.0% were of three ASCII characters;

4. 14.5% were strings of four alphanumerics;

5. 21.4% were five letters, all uppercase or all lowercase;

6. 18.3% were six letters, all lowercase;

Of the 3,289 passwords collected, 86% fell into one of the above classes. In addition, 15% of the passwords appeared in dictionaries, name lists and the like (Morris and Thompson, 1979). This makes user-selected password more vulnerable to being guessed by simply conducting a word search on a dictionary or word list in a word

processing software package. The lower half of the distribution it can be seen that users' passwords were typically made of normal alphabetic characters. This inference provides support to the premise that users typically choose passwords that are common or are some type of meaningful detail to them.

User-selected passwords have the advantage of being simple and meaningful. The disadvantage is that they are frequently based on trivial association and can be guessed by outsiders (Ahituv et al., 1987). This thesis takes a closer look at the characteristics of user generated passwords, password selection criteria and variables that may influence selection memorability and predictability.

3. Manufacturer-Generated Passwords

Manufacturers typically embed or hard-code passwords into programs. These embedded passwords serve as example passwords and are published in system documentation (Wood, 1983). Example passwords are intended to be temporary until a user selects a replacement. If a user does not remove an example password, it may become a source of vulnerability.

Another type of manufacturer's password is that used by field representatives and technicians. These passwords typically take the forms of test passwords and system passwords (Barton and Barton, 1984). They serve as a quick method by which technicians can gain access for maintenance and repairs. Knowledge of these passwords may allow unauthorized users to penetrate a security system.

4. Classification by Use

Passwords are also classified by their use: primary, secondary and duress. Primary passwords are used to gain access to an initial set of resources (Menkus, 1988).

Such as access to a building, a terminal room or terminal login. Secondary passwords are used as supplements to gain access to a subset of resources (Menkus, 1988). This includes access to classified databases, sensitive files and system level application controllers. Duress passwords are used when a user is under duress from a potential intruder. If a user is held at gunpoint, a duress password, instead of the usual password, can be entered to cause the system to simulate a crash, to alert the police or to take some other action unbeknownst to the person applying the duress (Wood, 1983). In addition, passwords can be dynamic in the sense of requiring a different password at each log-in (Avarne, 1988).

C. CONSTRUCTION OF PASSWORDS

The success of passwords as a security mechanism is related directly to good construction. Three criteria govern good construction: length, character set and memorability.

1. Length

The longer the password, the more difficult it is to guess it and therefore the more secure it is (Wood, 1983). Passwords are commonly constructed of six to eight characters. This length is popular for two reasons. First, six to eight characters are sufficient to guard against a "brute-force" attack (Wood, 1983). Second, memory aids are commonly required for recall of passwords of more than eight characters (Menkus, 1988). The elimination of memory aids decreases the probability that passwords will be committed to paper.

The minimum length of a password determines the lower bound of security (Menkus, 1988). Fisher (1983) suggests that the minimum length should be a set of

characters that would yield at least one million possible combinations. The following sets meet this minimum constraint: six decimal digits, e.g., 195863; five hexadecimal characters, e.g., 1D6FC; five alphabetic characters, e.g., AZHWO or four alphanumeric characters, e.g., HW39 (Fisher, 1984). A consideration in selecting a minimum length is that intruders will be attracted to trying all possible combinations in an exhaustive or brute-force attack. In an exhaustive attack, an intruder will need only try 40% of the possibilities to break a password (Menkus, 1988). A password composed of three numeric characters yields one thousand possibilities. A computer programmed to try each of the possible combinations will likely break the password in little time. Doubling length increases the effort required by orders of magnitude (Menkus, 1988). If three numerics were increased to six numerics, the combinations increase from one thousand to one million.

The design of the length of passwords should also consider whether a system will allow a user to construct a password that is shorter than the maximum. For example, if a password is designed to be eight numeric characters, will a system allow a user to use only four characters? Most systems will enter trailing blanks in the unfilled spaces (Menkus, 1988). A common ploy is for a potential penetrator to concentrate on trailing blanks first (Menkus, 1988). Elimination of the blanks reduces the total combinations that an intruder must attempt. By eliminating four trailing blanks, an intruder reduces the work factor from one hundred million to ten thousand possibilities.

2.   Character Set

The set of characters coupled with the number of characters determines the effectiveness of passwords. The ideal password is composed of random characters, such

as "k&)8[" (Barton and Barton, 1984). While random characters are more secure, they are seldom pronounceable. When a password is pronounceable, users will be better able to remember it (Kurzban, 1983). The addition of vowels increases pronounceability. However, the resulting password will be more vulnerable to attack (Kurzban, 1983). For example, if vowels are inserted into the string CTWLK, it becomes CATWALK.

3. Character Set vs. Length

As stated earlier the longer the password the more secure it is, coupling length with character set determines a password's effectiveness as an access control mechanism. Character set randomness is a key is a to an effective password. The greater the length and the more random the characters are in the password the lesser the predictability and the greater the time required for an intruder to penetrate a secure system using passwords as an access control mechanism. Morris and Thompson (1979) calculated the estimated time required to test all possible character strings of n length chosen from various sets of characters. A PDP-11/70 was the test instrument.

TABLE 3-1 Morris and Thompson's Table:

	26 lower case letters	36 lower case letters and digits	62 alphanumeric characters	95 printable characters	all 128 ASCII characters
n					
1	30msec	40msec	80msec	120msec	160msec
2	800msec	2sec	5sec	11sec	20sec
3	22sec	58sec	5min	17min	44min
4	10min	35min	5hrs	28hrs	93hrs
5	4hrs	21hrs	318hrs	112days	500days
6	107hrs	760hrs	2.2yrs	29yrs	174yrs

As be seen by the table, a password of considerable length and randomness provides a better degree of access control for information system security. The longer it takes a potential intruder to penetrate a system the greater the chances of detection and prevention (Morris and Thompson, 1979).

4.   **Memorability**

The ability to remember and recall passwords is of paramount importance in their construction. Most users require memory aids to help recall (Menkus, 1988). If a memory aid means writing the password on paper, a basic tenet of password security has been violated. A password committed to paper has changed from something known to something possessed (Porter, 1984). An intruder's work switches from guessing to searching.

An appeal to long term memory has been divided into two classifications of memory: semantic and episodic. These two classes form the basis for three approaches to enhancing the memorability of passwords: semantic, episodic and environmental (Barton and Barton, 1984).

Semantic memory uses information closely related to language use. Passwords using this approach are derived from well-known character strings, such as nursery rhymes. Nursery rhymes and similar strings are easily recalled, thereby eliminating the need for memory aids. For example, "Jack and Jill went up the hill" is a well known line from a childhood poem. In addition, these character strings are not related to a user's lifestyle. Once identified, the string can be used with a hashing

23

routine or a transform procedure to produce a phoneme, word or phrase that is actually the password.

Episodic memory relies directly on individual, personal experience. To a large degree, this experience will be unshared. Provided the user avoids the obvious references to experience, such as birthday dates and children's names, this type of memory is recommended for password systems. Transform procedures can operate in conjunction with episodic memory to produce passwords.

Environmental clues trigger the recall of passwords. A picture on the office wall or a room number can serve as the basis of a character string. If a user's terminal is located in a room that is painted green, "green walls" could serve as an initial character string. If a user's office is in room 821 at 1275 Sams Street, 8211275 could serve as an environmental trigger for a password. This string could then be manipulated by a transform procedure to produce the actual password. In the above example of 8211275, a transform procedure could take the even digits of 822 and add that result back to the initial room number to come up with the final password; i.e., 822 plus 821 equals 1643. In this example, 1643 is the password triggered by the environmental clue of the room number 821.

5.   Transform Procedures

Character strings produced by any of the three methods above can be coupled with transform procedures. A transform procedure manipulates a string to produce a user-recognizable and memorable password (Barton and Barton, 1984). Effective transform procedures are evaluated on the following criteria: ability to achieve a high degree of congeniality; i.e., easy to remember and to execute; ability to produce

24

structured passwords that can be recreated which helps error discovery and ability to produce passwords resistant to guessing and systematic trials. Common transform techniques are excerption and substitution. In excerption, a designated number of characters are excerpted based on their position within a string. The excerpted characters form the actual password. Substitution can also be used. Common substitution practices include the substitution of preceding or succeeding characters. The resulting string of substituted characters constitutes the password (Barton and Barton, 1984).

6. Mnemonics

Closely related to transform procedures are mnemonics. The phonetic sounding of a character string may yield an expression that is pronounceable and memorable (Barton and Barton, 1984). For example, the character string FRGTFL could be phonetically sounded as FOR-GET-FUL. While FRGTFL is the password, the phrase FOR-GET-FUL is the mnemonic that causes the password to be memorable. Other ways of avoiding memory aids are: inverting the order of characters, converting alphabetic characters to their numeric equivalents, shifting characters one or two positions and creating acronyms from initial letters of a meaningful phrase (Menkus, 1988).

7. Summary

Good formulation produces passwords that are distanced enough in form from ordinary experience to make compromise unlikely (Barton and Barton, 1984). Whether produced by semantic, episodic or environmental methodologies, passwords should be

25

evaluated for effectiveness. Ahituv et al. (1987) propose the following evaluation criteria:

1. should be easily memorized;

2. should be hard to guess through association;

3. should be easy to enter into the computer;

4. should not be able to be used if expired;

5. should be resistant to attack by spoofing or trojan horses;

6. should be tested;

7. should not take a long time to implement;

8. should not be cost prohibitive.

D. VULNERABILITIES OF PASSWORDS

The use of passwords as a security mechanism is a much debated topic. Opinions on effectiveness range from the criticism of they offer little resistance to a serious attack (Avarne, 1988) or their use is rarely well managed (Menkus, 1988) to the praise of the most cost-effective approach to human user authentication (Wood, 1983). Menkus (1988) makes the comparison of a password to a conventional lock; it keeps out only honest people.

Traditional passwords have three weaknesses: they can often be guessed, they are entered in the clear where they can be observed and they are used more than once (Avarne, 1988). These weaknesses are further supplemented by Ahituv et al. (1987): passwords are normally stored in tables in an operating system which itself is subject to compromise and spoof routines. Spoof routines, explained below, can be used during a log-in procedure to capture passwords from an unsuspecting user.

Eight methods of finding out a password have been identified: guessing, reading, hash tables, eavesdropping, intercept, signal radiation, spoofing and terminal buffers (Avarne, 1988).

1. Guessing

Users commonly use names, telephone numbers and other trivial but memorable data as passwords. Guessing entails repeated trials based on knowledge about a targeted user. To prevent guessing, systems may be equipped with counter programs that allow only a certain number of unsuccessful attempts before freezing out a would-be user. Such systems can still be penetrated through the intruder attempting one less than the maximum allowable attempts each day.

2. Reading

Passwords committed to paper are usually looked up just before a log-in. People nearby may see the location of a written password. Systems requiring frequent changes of passwords may increase the likelihood of users writing them down. In addition, frequent changes in passwords may be circumvented by re-entering an identical password or alternating between two passwords.

3. Hash Tables

Hash tables may lead to a false sense of security. An intruder needs only to know a hashed result of a password. Any character string that yields the same hashed result will suffice.

4. Eavesdropping

Most computer terminals do not echo a password back to the screen. Nonetheless, a person nearby may observe a sequence of keystrokes. Even listening to the number of keystrokes yields the length of a password.

5. Intercept

The proliferation of networks is a rich area for exploitation. Tapping into a line between a terminal and a host can give direct access to an intruder.

6. Signal Radiation

All electronic equipment, unless Tempest certified, emits radiomagnetic signals. These signals can be monitored and intercepted. Each keystroke emits a unique signal that can be correlated to give a direct interception of transmissions.

7. Spoofing

Penetrators develop programs that emulate terminal log-in procedures. A valid user enters a password not knowing that a spoof program is receiving the data instead of the computer. At the end of a log-in procedure, the computer gives an error message. The user assumes that a error has been made in keying in the information and re-enters the password. On the second try, the log-in is successful. Unbeknownst to an authorized user, an intruder now has a valid password and can enter the system at will.

8. Terminal Buffers

Passwords are written into a buffer from which the security program can read the entry. If a buffer is of large size or if system usage is low, a password may stay resident in a buffer for an indefinite time. An intruder monitoring a buffer may be able to read its passwords that are still resident.

E. MANAGING A PASSWORD SECURITY SYSTEM

How well a password system works depends primarily on how well the system is managed. Effective password system management involves proper system administration, implemmentation, protection and periodic evaluation.

1. Administration of Passwords

Password systems require maintenance. Akin to logical fences, passwords systems require periodic maintenance (Wood, 1983). In large systems, security may be in the hands of a full-time security manager. In smaller systems, security is likely to be part of a system administrator's job.

A security manager is responsible for maintaining and modifying a computer system's security. As well as duties related to passwords, a manager is responsible for physical security and disaster recovery. Monitoring a system for evidence of tampering and proper password use are a security manager's primary duties.

User education in security matters is also a concern of a security manager (Wood, 1983). Users have certain responsibilities when using the system and should be aware of the consequences of inappropriate actions (Panns and Herschberg, 1987). Education will make users aware of how a password system can protect their information from unauthorized access. At the same time, educated users will be aware

of how the design and protection of passwords can enhance overall system security. Help with developing passwords should be available on-line. Technical information about length, type of characters and ranges should be accessible to users (Barton and Barton, 1984).

2. Password System Implementation

Wood (1983) asserts that a password security system is successful if it meets the following criteria:

1. passwords are not visible when typed;

2. an alarm is generated if successive log-in attempts exceed a maximum, usually three;

3. a password storage table is encrypted and is not reversible;

4. passwords travelling over networks are encrypted;

5. provision is made for a special password to indicate a user is under duress and is being forced to log-in;

6. error messages are limited to a single message that does not indicate which step in the log-in process was wrong;

7. a password routine is segregated from the resource that it protects;

8. re-verification of a password is required if a session exceeds a time limit;

9. automatic log-off occurs if no activity takes place after a prescribed time period.

Successful implementation of system-generated passwords should include provisions for the secure distribution of passwords. Two common distribution methods are (1) conventional mail using double envelopes or specially designed envelopes that mask a password and (2) network transmission using encryption (Menkus, 1988). A user-selected password system eliminates the need for a password distribution system (Spender, 1987).

A password security system requires the commitment of top management. Information is a strategic resource. Lost or damaged information may have costly implications for an organization. Historically, hardware was the major cost of a computer system. In recent information systems, software is the major expense. Management often uses hardware values instead of the value of the information to base their security decisions (Wood, 1983).

3. Protection of Passwords

Successfully breaking a password may allow an unauthorized user total access to a computer system. In many systems, passwords are not only the first line of defense (Wood, 1983) they are the only line of defense. With the importance placed on passwords, security of passwords is a major concern. Passwords may be compromised by (Morris and Thompson, 1979; Pfleeger, 1989):

1. trying all possibilities;

2. trying all probable passwords;

3. trying passwords likely for a user;

4. searching for a system list of passwords;

5. asking a user.

Additional protection may be had through the use of encryption. Techniques include encryption of password tables stored in memory, use of one-time encrypted passwords and use of personal keys that are inserted after a plain text password is entered (Ahituv et al., 1987). One-way encryption increases the work needed to enter a system (Menkus, 1988). Encryption of password tables may be accomplished by the simple addition or subtraction of some constant (Menkus, 1988). Whichever encryption

method is used, care should be given to ensure that an encryption process does not expose encryption techniques used for other resources (Ahituv et al., 1987).

4. Improvements to Passwords and Password Systems

Cooper (1989), Morris and Thompson (1979) and Pfleeger (1989) identifies seven ways to improve the effectiveness of passwords as an access control mechanism:

1. use more than A - Z;

2. choose long passwords;

3. avoid actual names or words;

4. choose an unlikely password;

5. change the password regularly;

6. don't write it down;

7. don't tell anyone else.

Menkus (1988) identifies five ways to improve the performance of a password security system:

1. insist that an organization's policies are enforced;

2. prohibit storing of passwords in tables to speed up network connectivity;

3. penalize deliberate disclosure of passwords no matter how good the excuse;

4. require frequent changing and

5. insist that passwords be actually changed.

F. SUMMARY

Passwords can be an inexpensive, effective means to system security. The tradeoff between memorability (ease of use) and security will affect a user's environment. If a user's environment is unfriendly, a user will find ways of overcoming the difficulty and in turn, may compromise system security (Martin, 1973). A hostile environment is caused by an emphasis on security at the expense of password memorability (Barton and Barton, 1984).

IV. RESEARCH METHODOLOGY

A. INTRODUCTION

The purpose of this research is to identify the characteristics of user-generated passwords and determine if any causal relationships may exist between these characteristics and relevant variables (e.g., frequency of changing passwords, sensitivity of data and frequency of logging on). The study is based on statistical analysis of empirical data which were collected for this purpose at the Naval Postgraduate School (NPS). This chapter presents the data collection methodology, the statistical analysis strategy and the sample characteristics.

B. INSTRUMENTATION

In order to gather data about specific areas of interest with regards to user-generated password characteristics, a self-administered questionnaire was developed. A copy of this questionnaire is included in appendix A. The questionnaire focuses on three major issues: demographics, password characteristics and password memorability. The structure of the questionnaire is as follows:

1. Demographic Items

The first part of the questionnaire addressed a respondent's basic characteristics. Questions 1-3 asked for a respondent's age, sex and academic curriculum number (for students) or department number (for faculty/staff) respectively. The fourth question asked respondents if they use the NPS mainframe computer system.

If the respondent responded "yes", he or she proceeded to complete the remaining parts of the questionnaire.

2. Password Characteristics Data Items

The second part of the questionnaire addressed password structure and characteristics. Question number five asked a respondent to reveal the number of characters in his or her password. The questionnaire warned respondents not to reveal their specific password but only its characteristics. Respondents were then asked how they chose their password. There were five choices available:

1. meaningful detail such as a name, date or number;

2. a combination of meaningful details such as "Bill89" or "MaryJane#1";

3. a pronounceable password such as "2Bfree" or "eyewilllive";

4. random combination of characters;

5. other.

This question was followed by a similar question requesting the characteristics of the password. These could be alphabetic, numeric, alphanumeric or ASCII.

3. Data on Password Memorability and Computer Usage Characteristics

Whereas the previous questions specifically addressed the characteristics of a user-generated password, the remaining questions were concerned with a respondents computer usage and the memorability and predictability user-generated passwords. Question eight asked a respondent if he or she had difficulty remembering a password. If yes, question nine asked respondents to reveal if the password was written down. Expectations are that if a password is difficult to remember it will be written down.

If the password was written down, question ten asked a respondent to say where it was written down. If a password is difficult to remember but not written down expectations are that it will be forgotten. Once forgotten, a user no longer has access to the system unless a system administrator allows him or her to choose a new password or refreshes their memory by telling them their forgotten password. Question 11 asked respondents how often they change their password. The frequency of change could be due to expected intrusion into files, standard procedures or as a result of poor memory and not being able to remember the password. Most often if a password is easy to remember, guessing is easy. If guessing is easy, a password provides little access control. Question 12 asked respondents if they ever changed a password because they believed it had been guessed. This question is followed up by the question about what led them to believe it had been guessed.

Questions 14 and 15 asked a respondent to reveal the sensitivity and importance of his or her data on a scale of one to five. One was the lowest and five the highest degree of sensitivity and importance of the data. Data sensitivity refers to the degree to which problems would result if your data is disclosed. Data importance refers to value your data. Expectations are that a user would generate a password as secure as the importance and sensitivity of the data it controls access to.

Question 16 asked respondents from where do they normally work when using a computer system. There were four choices available:

1. private office at NPS;

2. home;

3. public terminal at NPS;

4. other.

Many cognitive psychologists propose that anything frequently used would be easy to remember. Frequency of use enhances memorability. Question 17 asked respondents to reveal the frequency by which they log on to the NPS mainframe system. Expectations are that with a high frequency of mainframe usage there would a high degree of password memorability. However, with a high frequency of logging on there is a hazard of guessing if a password is not frequently changed. Questions 18 and 19 asked respondents if they used computer systems other than the NPS mainframe system. If so, they were asked if they used the same password they used on the NPS mainframe.

C.   SAMPLE CHARACTERISTICS

The population consisted of students and faculty/staff members at NPS. There were roughly a total of 1600 students and 400 faculty and staff members in the population. A total of 2000 questionnaires were distributed using the schools internal mail system. 997 responded of which 208 were faculty/staff and 787 were students. The reason for selecting this particular population was both students and faculty, upon reporting to the school, are given computer mainframe accounts at the school's main computer center. These accounts can only be accessed with an account number provided by the computer center and a password the user creates. A two month time period was provided for respondents to return the questionnaire.

D. STATISTICAL ANALYSIS STRATEGY

1. Level of measurement

To relate or associate one concept with another the appropriate statistical or mathematical procedure must be selected. Level of measurement determines which statistical procedure is appropriate to test for relationships and associations. When one talks about the level of measurement, he or she usually mean the assigning of numbers to observations in such a way that the numbers are amenable to analysis by manipulation or operations according to certain rules (Siegel and Castellan, 1988). There are four levels of measurement that scientists consider in gathering data: nominal, ordinal, interval and dichotomous.

a. Nominal

When numbers or other symbols are used to identify the groups to which various objects belong, those numbers constitute a nominal or categorical scale (Siegel and Castellan, 1988). Numbers on football jerseys or automobile license plates comprise nominal scales. Word names such as Protestant, Catholic, Jew or Republican, Democrat, Libertarian, Socialist are examples of nominals. Variables being analyzed in this thesis constituting nominal levels of measurement are:

1. how password was chosen;

2. where respondent normally worked when using a computer system;

3. how password was guessed;

4. where password was written down;

5. characteristics of password.

38

b. Ordinal

It is frequently possible to order categories with respect to the degree to which they possess a certain characteristic (Blalock, 1979). If it is possible to have a complete rank ordering of classes, we have an ordinal scale. An example of this would be to classify families in the upper, upper-middle, middle, lower-middle and lower classes according to their socioeconomic status. Ordinal thesis variables are:

1. how vital is your data;

2. how sensitive is your data;

3. how often do you log on;

4. how often do you change your password;

c. Interval

When a scale has all the characteristics of an ordinal scale and when, in addition, the distances and differences between any two numbers on the scale have meaning, an interval scale level of measurement has been achieved (Siegel and Castellan, 1988). Temperature is an example of an interval scale level of measurement. The number of characters in the password was the only relevant interval scale variable in this thesis.

d. Dichotomous

A dichotomous level of measurement is a measurement of a variable with only two possible categories or values, such as yes or no and male or female (Nie, et al., 1975). The dichotomous thesis variables are :

1. was password difficult to remember;

2. was password written down;

3. was password guessed;

4. do you use non-NPS computer systems;

5. do you use the NPS mainframe;

6. do you use the NPS mainframe password on non-NPS system.

By classifying data by the levels of measurement the decision of which statistical test/model to use can be easily answered. The ability to manipulate data in order to test for relationships or associations between objects being observed or to obtain new information about the objects is dependent on the level of measurement.

2. Coding Technique

Each variable, based on its level of measurement, was assigned an identity number to facilitate data analysis. Nominal variables, such as how password was guessed, were assigned numerical values for each response type. For example, "meaningful detail" was assigned a value of one while "other" was assigned five. Ordinal variables assumed the same coding characteristics as nominal variables. For the nominal variable how often do you log on, the response "never" was assigned a value of one and "more than once a day" was assigned a nine. Interval variables assumed the values originally assigned to them by respondents. Dichotomous variables assume either a one or a zero. For the yes/no dichotomous variables, "yes" was assigned a one and "no" was assigned a zero. With the sex type dichotomous variable. a male respondent was assigned a value of one and female respondent was given a zero.

3. Testing

Based on the level of measurement, interpretable operations can be carried out on a given set of variables (Siegel and Castellan, 1988). Since the levels of measurement have been determined, the appropriate statistical test can now be performed to test for relationships and associations along with obtaining new information about the variables being tested.

Statistical testing can take the form of either descriptive or analytical statistical analysis. Descriptive analysis is concerned with summarizing and describing a given set of data. Examples of this would be arithmetic means test, standard deviations, modes and medians (Mansfield, 1983). Analytical statistical analysis is concerned with rational decision making under uncertainty, such as hypothesis testing, analysis of variance and regression analysis (Mansfield, 1983). Both types of analysis will be used in this thesis.

The most functional descriptive statistic to be used in the data analysis is the arithmetic mean. The means test determines the average observation in a given set of data. Also of interest is the distribution of the sample data amongst the various categories. Of more interest are the tests of associations or relations provided by the various analytical statistical tests available. With the level of measurement determined the appropriate statistical tests can be chosen. Blalock (1979) and Siegel and Castellan (1988) provide a decision matrix for determining the appropriate test for relation or association once the level of measurement has been determined. An applicable decision matrix derived using Blalock and Siegal's matrices is provided below. The axes if the

41

table reperesent variables in a relationship that are being tested for strength of association.

TABLE 4-1

LEVELS   OF   MEASUREMENT

	INTERVAL	ORDINAL	NOMINAL	DICHOTOMUS
INTERVAL	Pearson's R	----	ANOVA	T-TEST
ORDINAL	----	Spearman's R	Kruskal- - Wallis	Mann- - Whitney
NOMINAL	ANOVA	Kruskal- - Wallis	Cramer's V	Cramer's V
DICHOTOMUS	T-TEST	Mann- Whitney	Cramer's V	CHI-Square

LEVELS

OF

MEASURE-
MENT

As can be noted in the decision matrix, each variable based on its level of measurement requires a specific yet appropriate statistical test in order to test for relationships among the various variables.

4. Test Descriptions

a. Hypothesis testing

A hypothesis is a statement that can be tested with inferential statistical procedures (Porter and Hamm, 1986). Experimental hypotheses must be stated in terms of a null hypothesis and an alternative or research hypothesis. A null hypothesis is

hypothesis is usually a statement that there is no difference between populations being compared or that no association/relationship exist between the two. An alternative hypothesis, by default states that a difference or association does exist between the populations. An alternative hypothesis is "accepted" when a null hypothesis is rejected. Testing tries to reject the null hypothesis. Rejection occurs when a tabulated critical value, different for each statistical test (depends on the particular test, degrees of freedom and in most cases population size), is less than a test value found by carrying out computations unique to a particular set of variables. A level of significance is established for each statistical test. This level of significance value is the probability of rejecting a null hypothesis when actually it is true (Porter and Hamm, 1986). Typically, the level of significance is set at either .01 or .05. At an .01 level of significance, there is a 1% chance that the null hypothesis is true and rejecting it is a mistake. If an analytical test's computed significance value is less than the level of significance established for the test, then the findings from the test are said to be significant. The null hypothesis is therefore rejected.

b.   *Analysis of Variance (ANOVA) Test*

This is a parametric test used to test for associations between nominal and interval variables. Such a test produces a F-statistic and a significance value. This F-value compares the variability <u>between</u> the groups being tested to the variability <u>within</u> the groups. When compared to a tabulated critical value, this value determines whether a null hypothesis can or cannot be rejected. The significance value determines if the F-value could have been obtained by chance alone in most cases of a hypothetical large number of tests.

c. Chi-Square Test

The Chi-square test of homogeneity is used to analyze qualitative data from two dichotomus populations. It produces a Chi-square value and a significance value. The Chi-square value, when compared to a critical value extracted from the Chi-square table, determines if a null hypothesis can be rejected. The higher the Chi-square value the greater the degree of association between the two dichotomus variables.

d. Cramer's V Test

This a test of measurement of the degree of association or relation between nominal variables and nominal vs dichotomus variables. It produces a Cramer's coefficient of association and a significance value. The closer the coefficient is to unity, the greater is the degree of association between the two variables.

e. Kruskal-Wallis Test

A Kruskal-Wallis test determines whether an observed difference between groups is due to sampling error or treatment effect (Porter and Hamm, 1986). Similar to the ANOVA test, it produces an F-value and a significance value.

f. Mann-Whitney Test

This is used when testing ordinal data in rank order against dichotomus data. It is frequently used when the T-test cannot be used. The test produces a Mann-Whitney U value and a significance value. The Mann-Whitney U value, when compared to a critical value, determines whether to reject a null hypothesis that no association exists between the variables. The higher the U-value, the greater the degree of association.

g. *Spearman's R*

A Spearman's R test is used when both variables are ordinal. The resulting test values are a correlation coefficient and a significance value. the closer the coefficient value is to unity the greater the degree of association.

h. *T-Test*

A T-test is used to assess the difference between the arithmetic means of scores from two independent groups. It produces a T-value and a significance value. If a computed T-value is greater than a tabulated critical value, a null hypothesis of no association exist can be rejected.

V. DESCRIPTIVE FINDINGS

A. DEMOGRAPHIC FINDINGS

Of the 997 respondents, 903 were male, 92 were female while only two were missing data. The average age of respondents was 34, in a range from 23 to 76.

As stated earlier, of the 997 respondents, 208 were from faculty and staff while 789 were from students. The distribution of respondents from the various curricula was fairly evenly distributed. The majority of the respondents were from the hard science/computer-use oriented curricula. These included operations research/analysis and the computer technology curricula.

B. PASSWORD CHARACTERISTICS FINDINGS

Figure 5-1 reflects the distribution of the number of characters in each respondents password. The mean number of characters was six in a range from one to 15. 13.8% of the respondents did not reveal the number of characters in their password.

FIGURE 5-1

Number of Characters	Number of Respondents	Percent
1	1	0.1
2	4	0.5
3	48	5.6
4	144	16.8
5	206	24.0
6	214	24.9
7	120	14.0
8	112	13.0
9	4	0.5
10 or more	6	0.6
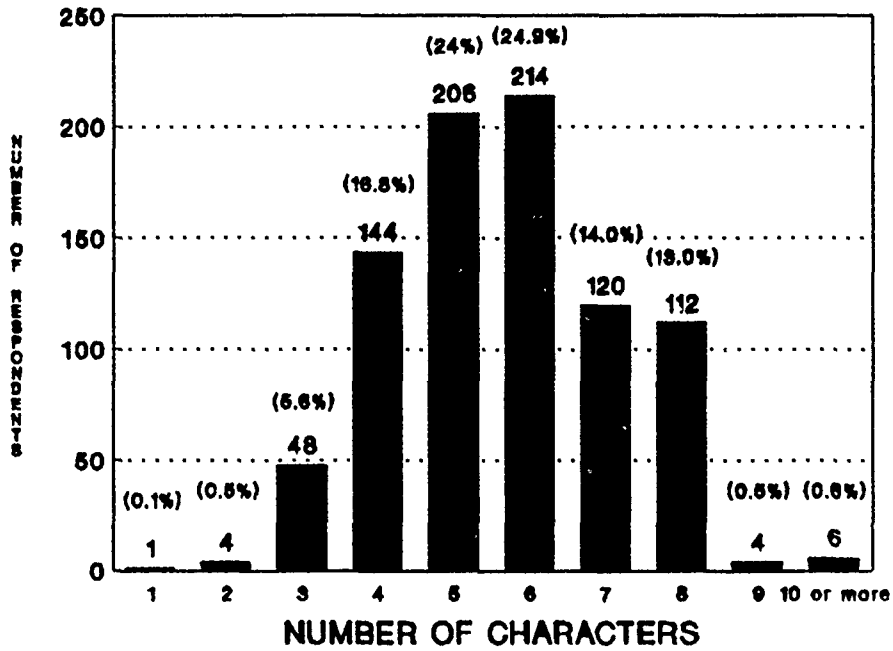
The distribution of how each respondent chose his or her password is shown in Figure 5-2 below. As expected, a plurality of the respondents (65.2%) chose their password from some type of meaningful detail. Meaningful detail was explained as some type of name, date or combination of the two to represent something meaningful to the respondent. 13.1% of the respondents did not answer this particular question.

FIGURE 5-2

Type of Choice	Number of Respondents	Percent
Meaningful Detail	561	65.2
Combination of Meaningful Detail	114	13.2
Pronounceable	47	5.5
Random Combination of Characters	19	2.2
Other	120	13.9

How Password was Chosen
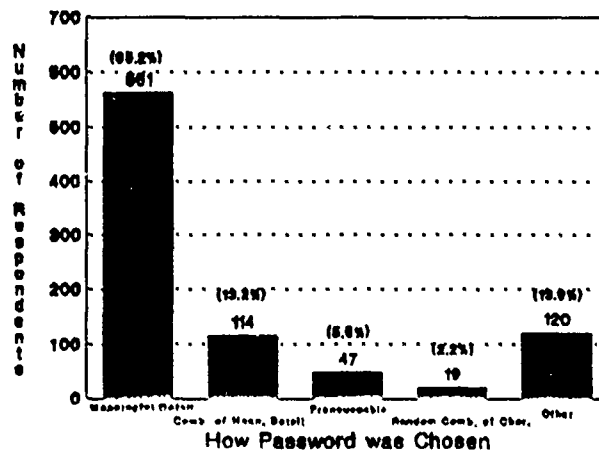
In revealing the characteristics of their passwords, 80.1% of the respondents who answered this question used passwords made up of alphabetic characters. The data in Figure 5-3 shows the distribution of the characteristics of each respondents password.

FIGURE 5-3

Characteristic	Number of Respondents	Percent
Alphabetic	689	80.1
Numeric	47	5.5
Alphanumeric	119	13.7
ASCII	6	0.7
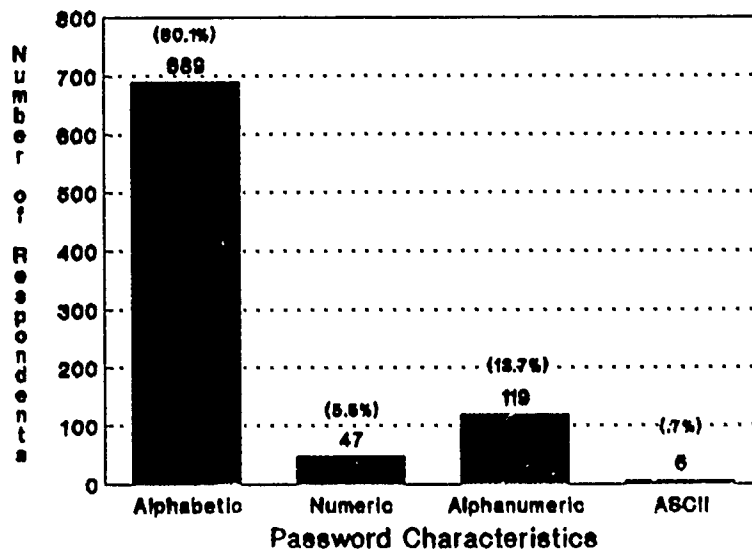
C. PASSWORD MEMORABILITY FINDINGS

How a password is chosen or the characteristics of a password should, in some way, affect its memorability. Relationships between these and other variables will be presented in Chapter VI. However, only 9.7% (83) of the respondents who answered this question, found it difficult to remember their password. 200 respondents found it necessary to write down their password. If a respondents password was written down Figure 5-4 shows where respondents said they wrote.

FIGURE 5-4

Place	Number of Respondents	Percent
Notebook	42	21.3
Wallet	83	42.1
Calendar	33	16.8
Desk/keyboard	4	2.0
Files	11	5.6
Other	24	12.2

Where Password was Written
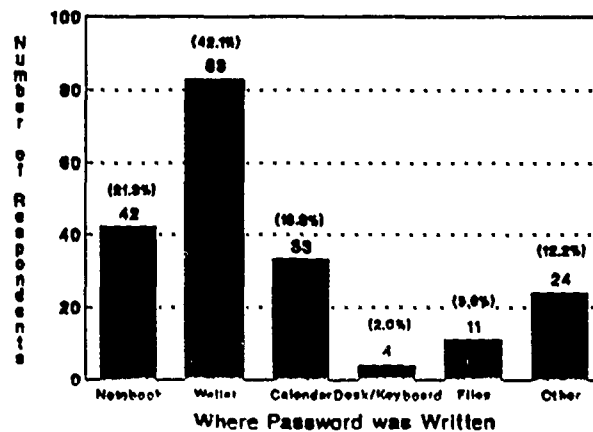
The frequency with which a password is changed may result from a password being difficult to remember, the suspicion that a password has been guessed, or security conscious procedure. The data in Figure 5-5 shows the distribution of the frequency with which passwords are changed by each respondent.

FIGURE 5-5

Frequency	Number of Respondents	Percent
Never	684	79.5
Less than once a year	128	14.9
Up to three times a year	33	3.8
4 - 6 times a year	8	0.9
Once a month	5	0.6
More than once a month	2	0.2

Frequency of Changing Password
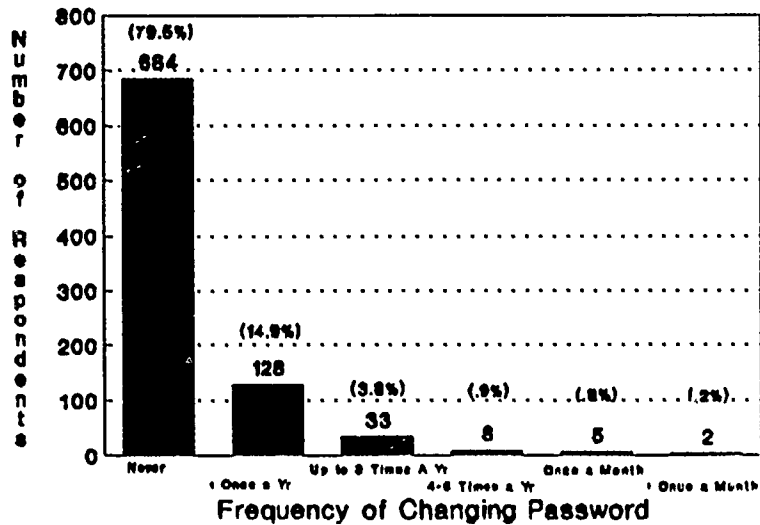
As can be noted, there were only 5.5% (48) of the respondents who changed their password with any degree of frequency.

Only 4.5% (38) of the respondents believed that their password had been guessed. The data in Figure 5-6 shows the distribution of the respondents reasons why they believed their password had been guessed.

FIGURE 5-6

Reason	Number of Respondents	Percent
Files altered	9	23.7
Told password	4	10.5
Unintentional Disclosure	10	26.3
Other	15	39.5

Why Believed Password had Been Guessed
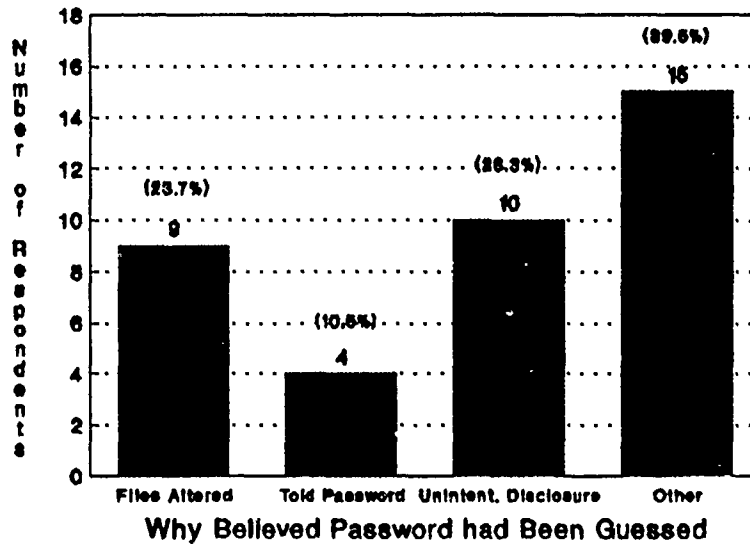
D. COMPUTER USAGE FINDINGS

The frequency with which a respondent logs on to the NPS mainframe was found to be fairly evenly distributed among the nine available choices. Figure 5-7 shows that roughly one fifth or about 19% of all respondents log onto the mainframe system several time a week.

FIGURE 5-7

Choice	Number of Respondents	Percent
Never	42	4.9
Annually	24	2.8
Quarterly	103	12.0
Once a month	104	12.1
Several times a month	149	17.4
Once a week	98	11.4
Several times a week	189	22.1
Once a day	83	9.7
Several times a day	64	7.5

Frequency of Logging On
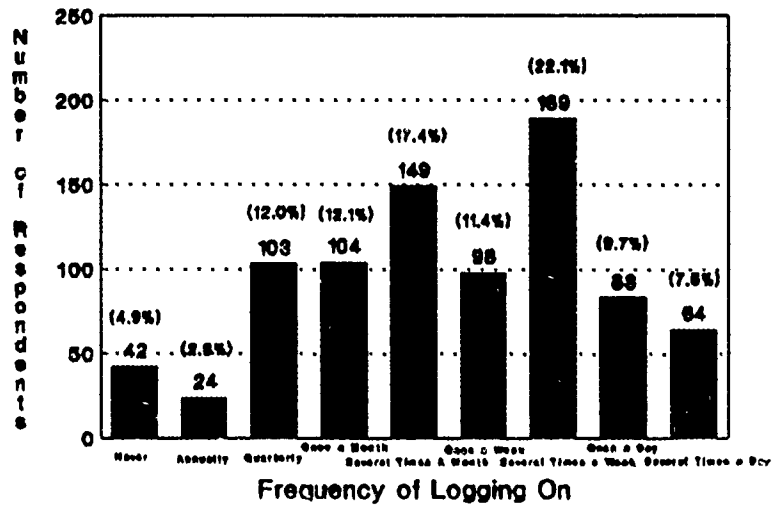
Tables 5-1 and 5-2 show the distribution of how each respondent viewed the sensitivity and importance of their data. With the lowest level of data sensitivity being one, it can be seen in Table 5-1 that most respondents (82%) viewed their data as not sensitive. However, the distribution for data importance was more evenly spread, with 15% of the respondents viewing their data as being vital.

TABLE 5-1

Level of Sensitivity	Number of Respondents	Percent
1	700	82.1
2	94	11.4
3	43	5.0
4	9	1.1
5	7	0.8

TABLE 5-2

Level of Importance	Number of Respondents	Percent
1	306	36.0
2	140	16.5
3	205	24.1
4	71	8.4
5	128	15.1

More than 50% of all respondents use NPS terminals when using a computer system. Table 5-3 shows that 18.6% (160) of the respondents use a NPS terminal in a private office while 57.9% (497) of the respondents use the public terminals at the NPS computer center or various labs.

TABLE 5-3

Place of work	Number of Respondents	Percent
Private NPS office	160	18.6
Home	185	21.6
Public terminal	497	57.9
Other	16	1.9

Twenty-seven percent (274) of all respondents use computer systems other than the NPS mainframe. Of those 274, less than half (104) used the same password used on the NPS mainframe on the non-NPS computer systems.

Analysis of statistical findings is presented in Chapter VI.

VI. ANALYTICAL FINDINGS

A. INTRODUCTION

You have just entered the main computer center of your organization. In a few moments you will be issued a computer account which requires the use of a password to access it. You choose a password you feel will meet computer center security requirements, provide you with access to your account, protect your resources and still be easy to use. What factors did you consider during the selection of your password? Will you write it down? Is it a secure password?

The analysis that follows looks closely at the data findings and attempts to relate what has been previously found to these findings. Of interest in this analysis is to see if password memorability and ease of guessing are associated with variables assessed by the self-administered survey. There are four areas of interest:

1) What variables are associated with a decision to write down a password.

2) What factors are related to how difficult a password is to remember.

3) The variables that are related to the ease of guessing a password.

4) How the sensitivity and importance of data are related to password selection.

All statistical computations were produced by the SPSS Statistical Package (Norusis, 1988). Relevant data printouts of the various statistical test computations can be found in Appendix B.

B. WRITING DOWN A PASSWORD

The variable "do you write down your password", a dichotomus variable, was tested with six other variables for associations. The assumptions of this analysis is that respondents will write down their password if:

1. it is difficult to remember;

2. it is not used often;

3. the characteristics of the password make it difficult to remember;

4. the password is selected poorly;

5. the password is changed often.

A null hypothesis was established for all tests. This null hypothesis stated that no association exists between the two variables being tested. The alternative hypothesis stated that some association exists. All tests were be conducted at the .01 level of significance. The findings of each test are presented in the table below. The horizontal axis represent the level of measurement assigned to each variable tested, the statistical test for association used for the two variables, the resulting test and significance values, the interpretation of these values and the location in the appendix of the data printouts associated with each test. The vertical axis represent the variables being tested for association against the test variable.

TABLE 6-1

Do You Write Down Your Password						
Variable (1)	Level of Measure- -ment	Test Conducted	Test Value (2)	Sign. Value	Interpre- -tation (3)	Refer to Appendix
Number	Interval	T-test	-.20	.839	NA - SI	B-1
Password	Nominal	Crmr's V	.1194	.0065	WA - SS	B-2
Chosen	Nominal	Crmr's V	.0875	.1584	NA - SI	B-3
Change	Ordinal	Mann-Whit	65972	.9899	SA - SI	B-4
Remember	Dichotomus	CHI-sqr	38.45	.0000	MA - SS	B-5
Logging	Ordinal	Mann-Whit	49783	.0000	SA - SS	B-6

LEGEND

1. Variables: Number - Number of characters in password.
   Password - Password characteristics.
   Chosen - How password was chosen.
   Change - How often password is changed.
   Remember - Do you have difficulty remembering password.
   Logging on - How often do you log on to a system.

2. Test values: The CHI-Square test has a critical value of 6.64.
   The T-test has a critical value of 1.645.

3. Interpretation: NA - No association exists between the two variables.
   WA - A weak association exists.
   MA - A medium association exists.
   SA - A strong association exists.
   SI - A statistically insignificant finding.
   SS - A statistically significant finding.

System-generated passwords typically consist of pseudo-random characters (Wood, 1983; Menkus, 1988). They, therefore, tend to be complicated, difficult to remember and unpopular with users (Wood, 1983). If a password is not easy to remember then users tend to write it down (Avarne. 1988). 83 respondents found it difficult to remember their password. However. 200 respondents felt it was necessary to write

58

down their password. Users who perceive they will not be using the computer system on frequent basis may chose to write down their password for future reference. Users may write down a password simply out of habit. Or users may write down a password because frequent change requirements are too demanding for their mental capacity or desire to remember. More change increases the likelihood a password will be forgotten.

The number of characters in a password was found to have a statistically insignificant non-association with writting down a password. This finding is not surprising; the number of characters in a password is, however, expected to affect memorability. It follows that if password is difficult to remember it is written down (Avarne, 1988).

Password characteristics was found to have a weak, statistically significant association with writting down a password. Here expectations also were that password characteristics and how a password was chosen affects password memorability which may lead to writting down a password.

How a password was chosen was found to have a statistically insignificant non-association with whether a password is written down.

The memorability of a password influences password changes. If it is difficult to remember then it is changed, written down or both. Data findings revealed a strong association between the frequency of changing a password and whether it is written down. However, the finding was not statistically significant.

A medium, statistically significant association was found between password memorability and whether it is written down. This finding strongly reinforces previous research. If a password is difficult to remember it is written down (Avarne, 1988).

A strong association was found between the frequency of logging on to a system and whether a password is written down. As mentioned earlier, the higher the use of a password the less likely it is to be forgotten. If it is not forgotten the need to write it down is reduced. Once a user writes down a password, he or she is inclined to put it in an insecure place (Spender, 1987). Of the respondents who wrote down their passwords, 21.3% wrote it in a notebook, 42.1% stored it in their wallets, 16.8% on a calendar and 12.2% stored the password in some sort of file. Once a password is written down it is no longer something known, it becomes something possessed (Porter, 1982). Knowledge of the place where it is stored becomes something known. Searching through a user's notebook, desk or diary is a good way of discovering a password (Avarne, 1988).

C. DIFFICULTY REMEMBERING A PASSWORD

Testing for association between variables that may relate to password memorability was of particular interest for this analysis. An assumption was made that if a password is difficult to remember then it is written down. A password that is long is expected to be more difficult to remember than one that is short. Password characteristics and how the password was chosen are expected to be associated with password memorability. Also, expectations are that if a password is frequently used it is less likely to be forgotten. Considerations for how often a password is changed and whether the password is used on other systems also is of interest. Difficulty

60

remembering a password is a dichotomus variable. A null hypothesis was established for testing for association between difficulty remembering your password and the other variables mentioned earlier. The null hypothesis states that no association exists. All testing was conducted at the .01 level of significance. The table below shows the results of the tests for association.

TABLE 6-2

Do You Have Difficulty Remembering a Password						
Variable (1)	Level of Measure- -ment	Test Conducted	Test Value (2)	Sign. Value	Interpre- -tation (3)	Refer to Appendix
Write	Dichotomus	CHI-sqr	38.45	.0000	MA - SS	B-5
Number	Interval	T-test	-.38	.706	NA - SI	B-1
Password	Nominal	Crmr's V	.1131	.0110	WA - SS	B-7
Chosen	Nominal	Crmr's V	.1221	.0121	WA - SS	B-8
Logging	Ordinal	Mann-Whit	26259	.0214	SA - SS	B-6
Change	Ordinal	Mann-Whit	25363	.0000	SA - SS	B-4
Passsame	Dichotomus	CHI-sqr	1.475	.2245	NA - SI	B-9

LEGEND

1. Variables: Number - Number of characters in password.
   Password - Password characteristics.
   Chosen - How password was chosen.
   Change - How often password is changed.
   Write - Do you write down your password.
   Logging on - How often do you log on to a system.
   Passsame - Do you use the same NPS password on other systems.

2. Test values: The CHI-Square tests have critical a value of 6.64.
   The T-test has a critical value of 1.645.

3. Interpretation: NA - No association exists between the two variables.
   WA - A weak association exists.
   MA - A medium association exists.
   SA - A strong association exists.
   SI - A statistically insignificant finding.
   SS - A statistically significant finding.

It has long been accepted that people can remember expressions of about seven characters in length (Menkus, 1988) and typically remember expressions that have particular meaning to them. Password length, characteristics, frequency of use and change and how it was chosen are believed to influence password memorability Menkus (1988).

As presented in the previous section password memorability and whether a password is written down were found to have a significant, medium association. If a password is difficult to remember it is forgotten if it is not written down.

Barton and Barton (1984) and Menkus (1988) suggests that the ability to recall a password tends to decrease as length increases. The average length of respondents passwords in this study was six characters. The ideal length is six to eight characters (Menkus, 1988). Only 83 respondents found it difficult to remember their password. Only 10 respondents had passwords greater than eight characters. This indicates that many of the respondents who had difficulty remembering their password, did so for reasons other than length. The hypothesis test revealed a statistically insignificant non-association exist between the number of characters in a password and difficulty in remembering it. Other factors have a greater influence on password memorability for this particular population.

The characteristics of a password or how it was chosen could have a greater effect than password length. Users who choose their own password are more likely to remember them (Wood, 1983). Users select from a simple domain of things meaningful to them, something from episodic memory (Menkus, 1988; Wood, 1983). Of the 997 respondents here. 675 chose passwords from meaningful details. 807

constructed passwords of alphabetic or alphanumeric characters. Weak but significant associations were found between password memorability and password characteristics as well as how a password was chosen. This can be interpreted as a positive relationship. As previous research revealed, an alphanumeric password chosen from meaningful detail is more easily remembered than passwords generated from pseudo-random combinations (Woods, 1983).

The frequency of use of a password should be related to password memorability. 687 respondents used their password with a frequency of once a month or more. The association between difficulty remembering and frequency of logging on was found to be strong but statistically insignificant. This supports the assumption that log on frequency is related to password memorability.

Data analysis findings revealed a strong association between the frequency of changing a password and password memorability. This supports previous research. Frequently changing a password diminishes password memorability. Frequent change increases the likelihood of a password being forgotten (Spender, 1987). Only 176 respondents changed their password with any degree of frequency.

D. GUESSING A PASSWORD

Respondents were asked whether their password had been guessed. If the answer to this question was "yes" , immediately following was the question of "what led them to believe it had been guessed? " 23% of the respondents who felt their password had been guessed were led to that belief because their files had been altered. 26% believed their password had been guessed due to unintentional disclosure; 10% intentional disclosure and 39.5% some other action led them to believe that their password had

been guessed. The objective that should be achieved when selecting a password is to maximize ease of use (memorability) and minimize predictability. Of interest here are associations between variables believed to be related to password predictability. " Was your password guessed " is a dichotomus variable. It will be tested against how often a password is changed, how a password was chosen, the frequency of logging on to a system, password characteristics, whether a password was written down, data importance and the number of characters in a password. Previous studies provides a basis for assumming that password predictability is related to these variables. A null hypothesis was established stating that no association exists between predictability and these other factors. Table 6-3 shows the results of the statistical testing.

TABLE 6-3

Was Your Password Guessed						
Variable (1)	Level of Measure- -ment	Test Conducted	Test Value (2)	Sign. Value	Interpre- -tation (3)	Refer to Appendix
Number	Interval	T-test	-1.27	.204	NA - SI	B-1
Write	Dichotomus	CHI-sqr	.5280	.4674	NA - SI	B-10
Password	Nominal	Crmr's V	.1445	.0004	WA - SS	B-11
Chosen	Nominal	Crmr's V	.0935	.1145	NA - SI	B-12
Logging	Ordinal	Mann-Whit	985.5	.4677	WA - SI	B-4
Change	Ordinal	Mann-Whit	64125	.0000	SA - SS	B-6
Work	Nominal	Crmr's V	.2138	.0000	MA - SS	B-13

Password compromises have resulted from information on computer bulletin boards, from guesses about personal vitae, environmental cues and from systematic intrusions (Barton and Barton, 1984). Experience shows that people prefer passwords that are easy to remember (Avarne, 1988). Easy to remember passwords are usually some form of meaningful detail, simply structured and highly predictable. Guessing a password is expected to be influenced by password characteristics, frequency of use, whether the password was written down, how often it is changed, frequency of logging on or from where a user normally works.

Only 38 respondents felt their password had been guessed. 23% of them felt their password had been guessed because their files had been altered; 26% because of unintentional disclosure, 10% intentionally revealed their password and 39.5% of the respondents who's password had been guessed felt it had been guessed for some other reason. 63.2% respondents who believed their password had been guessed had no idea how it was guessed.

Testing for association between the number of characters in a password and password predictability revealed a statistically insignificant non-association exists. Morris and Thompson (1979), in their study, revealed the shorter a password the less time require by an intruder, using a brute force attack, to reveal the password. Whether a password is written down was not associated with predictability. This finding is not in standing with previous research that suggests once a password is written down it becomes something possessed that can be lost or stolen.

Password characteristics was found to have a weak statistically significant association with predictability. This supports earlier research. Relatively short passwords chosen from some form of meaningful detail and consisting of alphanumerics make guessing easy. Because of this, Morris and Thompson's (1979) study revealed that an intruder conducting a dictionary search alone, would require only five minutes to reveal about one-third of the 3,289 passwords collected.

Tests for association between how a password was chosen and password predictability revealed that how a password was chosen had no association with guessing. This contradicts previous research. Refering back to the Morris and Thompson table presented in Chapter III, they revealed that passwords consisting of simple letters/numerics required less time for prediction than those made of ASCII characters. The test for association between log on frequency and password predictability revealed a weak but statistically insignificant association. The frequency of logging on to a system is expected to increase predictability. When a user logs on to a system, someone can capture the password by watching keystrokes (Ahituv et al.,

1988). As stated earlier, 687 respondents logged on to a computer system more than once a month.

Most previous research strongly support the frequent changing of passwords to insure system security and reduce guessing. Wood (1983) asserts that passwords should be changed annually. Menkus (1988) suggests every 30 days. Although changing a password increases the level of system security, it hinders memorability. The trade-off of ease of use and security must be optimized. Only 176 respondents changed their password more than once a year. The test for association between frequency of change and password predictability revealed a significantly strong association. This association supports previous studies and the expectation that frequency of change decreases predictability.

Where a user normally works when using a system is also expected to influence predictability. Using a public terminal is more vulnerable than working in a private office or at home. Of the 997 respondents, 185 worked at home, 160 worked from a private office and 497 worked on public terminals. Those who worked at home were considerably less vulnerable than the others. There was a medium, significant association between work location and password predictability.

E. FINDINGS ON DATA SENSITIVITY AND DATA IMPORTANCE

Respondents were asked to rate the sensitivity and importance of their data on a scale of one to five. Data sensitivity refers to the degree to which embarrassment or problems would result from the disclosure of the data. Data importance can be seen as the value or utility of the data to the respondent. The importance of data is assumed here to increase the degree of security desired. This analysis examines the association of password characteristics with data importance and sensitivity. The null hypothesis is that no association exists. The following table shows the results of statistical testing.

67

TABLE 6-4

Level of Data Importance						
Variable (1)	Level of Measure- -ment	Test Conducted	Test Value (2)	Sign. Value	Interpre- -tation (3)	Refer to Appendix
Chosen	Nominal	Kruskal-W	12.98	.0114	NA - SS	B-14
Password	Nominal	Kruskal-W	8.073	.0889	NA - SI	B-15
Number	Interval	ANOVA	.430	.787	NA - SI	B-16
Work	Nominal	Kruskal-W	91.79	.0000	SA - SS	B-15
Write	Dichotomus	Mann-Whit	55157	.0020	SA - SS	B-17
Change	Ordinal	Spearman	.1916	.0000	WA - SS	B-18

Level of Data Sensitivity						
Variable	Level of Measure- -ment	Test Conducted	Test Value	Sign. Value	Interpre- -tation	Refer to Appendix
Chosen	Nominal	Kruskal-W	7.264	.1226	NA - SS	B-15
Password	Nominal	Kruskal-W	2.886	.5771	NA - SI	B-14
Number	Interval	ANOVA	1.388	.236	NA - SI	B-16
Work	Nominal	Kruskal-W	29.13	.0000	MA - SS	B-14
Write	Dichotomus	Mann-Whit	64272	.8915	SA - SI	B-17
Change	Ordinal	Spearman	.1544	.0000	WA - SS	B-18

LEGEND

1. Variables: Number - Number of characters in password.
   Password - Password characteristics.
   Chosen - How password was chosen.
   Change - How often password is changed.
   Write - Do you down your password.
   Work - From where do you normally work when using a system.

2. Test values: The Kruskal-Wallis tests have critical a value of 13.277.
   The ANOVA tests have critical values of 3.32.

3. Interpretation: NA - No association exists between the two variables
   WA - A weak association exists.
   MA - A medium association exists.
   SA - A strong association exists.
   SI - A statistically insignificant finding.
   SS - A statistically significant finding.

The level of security should be commensurate with the importance of the data it protects (Hoffman, 1977). Users are expected to choose passwords that will provide the degree of security commensurate with the data they protect. Respondents were asked to rate the importance and sensitivity of their data on a scale of one to five. 59 respondents felt their data was moderately to highly sensitive, while 474 respondents felt their data was moderately to highly important.

A secure password is one that is relatively long, made up of random alphanumerics, is easy to remember and difficult to guess. Once again, the trade-off of ease of use and security must be optimized. This is difficult to achieve. Data sensitivity and importance is expected to have some influence on how a password is chosen, the number of characters in the password, the characteristics of the password, whether the password is written down, how often is changed, and from where the user normally works when using a system.

There were no associations between data importance or sensitivity and how a password was chosen, password characteristics or the number of characters in a password. The non-association can be explained by understanding that most students, when issued a mainframe account, have little idea what they will be storing in their accounts or it's importance. This can be related to the finding that few respondents changed their password throughout their use of a mainframe account.

Data importance and work location were found to have a statistically strong associated. Typically, if a user is working on an important data file, the place to work is somewhere private, unexposed to typical vulnerabilities.

Whether a password is written down and the importance or sensitivity of the data were found to be strongly related. A security conscious user with important and sensitive data files will most likely not chose to write down his or her password for fear of it being lost. Once written down the degree of security is compromised.

There was a weak association between the frequency of changing a password and both data importance and sensitivity. This can be attributed to the premise that a security conscious user will frequently change his or her password in order to reduce the likelihood of it being guessed and increase the protection of his or her data files.

These findings indicate that several factors play an importance role in user-generated password selection, memorability and predictability.

VII. CONCLUSIONS AND RECOMMENDATIONS

A. PASSWORDS AS AN EFFECTIVE ACCESS CONTROL MECHANISM

This thesis pointed out that access control is required at various levels in order to obtain a required level of security. At each level a certain amount of user identification, authentication and authorization must be verified. Passwords were found to be an effective mechanism for such. Traditional passwords, however, have some inadequacies. Morris and Thompson's study revealed some of the inadequacies of user-generated passwords in the pre-personal computer era. Some of these inadequacies included passwords relatively short in character length and passwords made of some type of meaningful detail to the user making them easy to remember. Passwords that are easy to remember provide low levels of security. This thesis follows Morris and Thompson's research in identifying the characteristics of user-generated passwords but in personal computer era. The characteristics of user-generated passwords were identified along with how these characteristics affect password selection, memorability and predictability. A new area brought to light is how does the importance and sensitivity of a user's data affect password selection, memorability and predictability.

B. CHARACTERISTICS OF USER-GENERATED PASSWORDS

1. Pre-personal Computer Era Characteristics vs. Personal Computer Era Characteristics.

This thesis has shown that the characteristics of user-generated password in the personal computer era have not changed much from those characteristics in the pre-personal computer era identified by Morris and Thompson (1979). User-generated passwords of today still bear the characteristics of being made up of some type of meaningful detail to the user, relatively short in length, made of alphabetic/alphanumeric characters and typically written down on paper. In general, they remain easy to remember and simple in structure. However, what has changed is the users attitude toward computer security. The impetus of system security has made the common user more privy to computer security requirements and more receptive to organizational administrative and technical security controls/procedures.

2. Password Characteristics and Writing Down a Password.

Most users require memory aids to help recall (Menkus, 1988). The most common type of memory aide is writing the password on paper. This violates the basic tenet of password security. Typically, a password is written down if it is difficult to remember (Avarne, 1988). However, passwords are also written down simply out of habit, from the perception that the password will not be used often enough to remember or because of system change requirements are too demanding to remember each password. This research showed that password memorability affects whether a password is written down.

3. **Password Characteristics and Password Memorability.**

This research revealed that several password characteristics affect password memorability. The findings that support previous research were: password characteristics and how a password is chosen (meaningful detail, combination of meaningful detail, pronounceable, etc,.) affect password memorability, the frequency of changing a password, although increases the level of system security, hinders memorability; the frequency of logging on, may in many cases hinder security if the password is not changed, enhances memorability. Most noteworthy is the finding that password length was found not to have any affect on memorability. This can be attributed to the advent of pronounceable passwords (mnemonics) such as "2good2Btrue" and passphrases such as I Love Paris In The Springtime (phrase) - ILPITST (password) (Menkus, 1988; Barton and Barton, 1984).

4. **Password Characteristics and Password Guessing.**

Results of this research show that password predictability is strongly affected by the frequency of changing a password. As previous research purports, the greater the frequency of change the greater the level of system security. Although previous research suggested that passwords made of meaningful detail, relatively short in length and simple in structure leads to ease of guessing the findings of this research did not support this. A noteworthy finding that surprisingly goes against previous research is writing down a password was found not to affect password predictability. Writing down a password violates the basic tenet of password security in that the password is something known, something secure but when it written down it becomes something possessed that can be lost, placed in an insecure place or stolen.

5. Password Characteristics and The Level of Data Importance and Sensitivity.

Although previous research revealed very little on this area of interest, this research shows that data importance and sensitivity does affect certain characteristics of user-generated passwords. Hoffman (1977) suggests that the level of security should be commensurate with the importance of the resources it protects. Although, many respondents for this research did note rate their data as very important nor sensitive, the few that did were expected to exercise sound password security principles for password selection and use. This study showed that how a password is chosen, the number of characters in a password and password characteristics (alphabetic, alphanumeric, ASCII, etc,.) were not affected by the level of data importance and sensitivity. This finding can be understood by noting that most respondents (students) upon reporting to NPS and being asked to choose a password have no idea what or the importance of what will be stored in their mainframe accounts. It can also be noted that many, in the onset, are not very computer security conscious. Data importance and sensitivity was found to strongly affect where a user will work when using a system. A security conscious user working on sensitive and importance data will typically work in a location that is private, less exposed to the threats of security. This research also revealed that the frequency of changing a password is affected by the level of data importance and sensitivity. A security conscious user will opt to changed his or her password more frequently if they are protecting data which is important and sensitive.

C. RECOMMENDATIONS

This study particularly shows that the characteristics of user-generated passwords in the personal computer era have not changed drastically from those characteristics in the pre-personal computer era. What has changed is the users awareness of the impetus of computer security.

This study also show that certain characteristics of user-generated passwords affect password selection, memorability, and predictability. Most importantly revealed was the level of data importance and sensitivity affect password selection and predictability.

Following the recommendations of Cooper (1989), Morris and Thompson (1979) and Pfleeger (1988), in order to improve the level of security/access control provided by passwords they sould be:

1. longer in length;

2. made of meaningful detail to aide in remembering;

3. greater mix of characters such as ASCII characters;

4. frequently changed;

5. not written down.

Although, passwords are still widely used, confidence in their capabilities in providing adequate security is decreasing. Applications of passwords as a security mechanism have not advanced as rapidly as information system technology. Because of this, the details of password systems applications and their effectiveness warrant further research.

Thesis Questionnaire - Computer Password Characteristics

Improving effective information system security is a continuing problem. Passwords are widely used to control access to information systems. The purpose of the questionnaire is to generate sample data on the characteristics of user generated passwords at the NPS. I do not want to know your password, only certain characteristics about it. The resulting data will be used to create a new form of passwords that are difficult to guess.

NOTE:

Even if you are not a computer user or do not use the computer frequently your response to this questionnaire will still provide us with important information.

PART A: Personal Information

1. Age : _____

2. Sex (circle one) : Male   Female

3. Curriculum (Students) : _____
      or
    Department (Faculty) : _____

PART B: Password Characteristics (Please do not reveal your password !!)

1. Do you use the NPS mainframe system (circle one) ?

   No

   Yes

   If no, please return this questionnaire anyway. Even if you do not use the NPS system, we appreciate completed returns to this survey.

   If yes, please continue.

2. How many characters are in your password ?   _____

3. How did you choose your password (circle one)?

    A. A meaningful detail. (e.g., name, date, street)

    B. A combination of meaningful details. (e.g., Bill1989, 4june63)

    C. A pronoucable password. (e.g., one4you, 2Bfree)

    D. A random combination of characters. (e.g., carS&, dUCk*?+)

    E. Other (please specify). _____

4. What are the characteristics of your password (circle one) ?

    A. Alphabetic (e.g., abdc, ERTIS) .

    B. Numeric (e.g., 1234, 5879).

    C. Alphanumeric (e.g., a34d, fo67YI).

    D. ASCII (e.g., cd!Yx, Ac1+t6).

5. Have you ever had difficulty remembering your passwords (circle one) ?

    No

    Yes

6. Very often, computer users find it convenient to write down their
password for one those unfortunate times when they forget it.
Do you also practice this (circle one) ?

    Yes

    No

If so, where do you write it down (users manual, calendar
book, notebook, keyboard, on something in your wallet) ?

    where _____

7. How often did/do you change your password (circle one) ?

    A. Never
    B. Less than once a year
    C. Up to three times a year
    D. Four to six times a year
    E. About once every month
    F. More than once a month

8. Have you ever changed your password because you felt it had been guessed by someone else(circle one) ?

        Yes

        No

If so, what led you to believe it had been guessed ?

_____

_____

9. On a scale of one to five, how sensitive are your data (what problems would result if revealed) (circle one) ?

```
  1        2        3        4        5
  |_____|_____|_____|_____|
 Non-             Moderately        Very
Sensitive         Sensitive       Sensitive
(nothing          (mildly      (embarrassing personally
to hide)        embarrassing)    or to the organization)
```
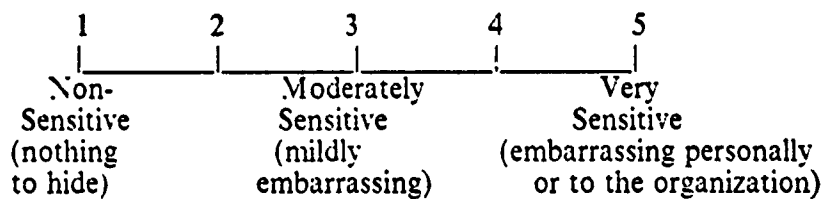
10. How important are your data (how vital are your data) (circle one) ?

```
  1        2        3        4        5
  |_____|_____|_____|_____|
 Non-             Moderately       Highly
Vital             Vital            Vital
(not important,               (thesis, research
would not miss,                  results)
life would go on)
```
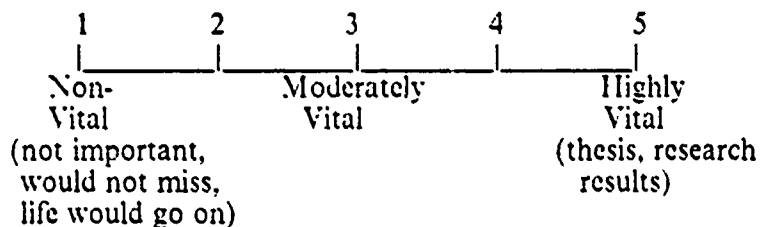
11. When using a computer system, from where do you normally work (circle one) ?

A. Private office at NPS

B. Home

C. Public terminal at NPS

D. Other (please specify) _____

12. How often do you log on to the NPS mainframe (or other NPS system) (circle one)?

    A. Never

    B. Annually

    C. Quarterly

    D. At least once a month

    E. Several times a month

    F. At least once a week

    G. Several times a week

    H. At least once a day

    I. Several times a day

13. Do you use any non-NPS computer systems which require the use of a password (circle one) ?

    Yes

    No

14. Do you use the same NPS password on non- NPS systems (circle one) ?
    Yes

    No

Please place completed questionnaire in the self-addressed envelope
provided and return as soon as possible.

Thank you for your cooperation,

D. A. Sawyer, LT USNR
SMC 2341

```
- - - - - - - - - - - - - - - - - - - - - - - - - - T - T E S T - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

GROUP 1 - DIFFREN EQ   0: NO
GROUP 2 - DIFFREN EQ   1: YES

Variable	Number of Cases	Mean	Standard Deviation	Standard Error	*	F Value	2-tail Prob.	*	t Value	Degrees of Freedom	2-tail Prob.	*	t Value	Degrees of Freedom	2-tail Prob.
									Pooled Variance estimate			**Separate Variance Estimate**			
PASSNUM NUMBER OF CHARACTERS					*			*				*			
GROUP 1	775	5.6800	1.545	.055	*			*				*			
					*	1.10	.594	*	-.38	856	.706	*	-.39	102.33	.696
GROUP 2	83	5.7470	1.472	.162	*			*				*			

```
- - - - - - - - - - - - - - - - - - - - - - - - - - T - T E S T - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

GROUP 1 - GUESSED EQ   0: NO
GROUP 2 - GUESSED EQ   1: YES

Variable	Number of Cases	Mean	Standard Deviation	Standard Error	*	F Value	2-tail Prob.	*	t Value	Degrees of Freedom	2-tail Prob.	*	t Value	Degrees of Freedom	2-tail Prob.
									Pooled Variance estimate			**Separate Variance Estimate**			
PASSNUM NUMBER OF CHARACTERS					*			*				*			
GROUP 1	812	5.6761	1.496	.053	*			*				*			
					*	2.20	.000	*	-1.27	848	.204	*	-.89	38.59	.379
GROUP 2	38	6.0000	2.218	.360	*			*				*			

```
- - - - - - - - - - - - - - - - - - - - - - - - - - T - T E S T - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

GROUP 1 - WRITTEN EQ   0: NO
GROUP 2 - WRITTEN EQ   1: YES

Variable	Number of Cases	Mean	Standard Deviation	Standard Error	*	F Value	2-tail Prob.	*	t Value	Degrees of Freedom	2-tail Prob.	*	t Value	Degrees of Freedom	2-tail Prob.
									Pooled Variance estimate			**Separate Variance Estimate**			
PASSNUM NUMBER OF CHARACTERS					*			*				*			
GROUP 1	659	5.6798	1.544	.060	*			*				*			
					*	1.04	.737	*	-.20	857	.839	*	-.21	334.62	.838
GROUP 2	200	5.7050	1.513	.107	*			*				*			

WRITTEN IS PASSWORD WRITTEN DOWN? by PASSCHAR PASSWORD CHARACTERISTICS

```
                 PASSCHAR                        Page 1 of 1
            Count I
            Row Pct IALPHABET NUMERIC  ALPHANUM ASCII
            Col Pct IIC                   ERIC            Row
            Std Res I    1  I    2  I    3  I    4  I Total
WRITTEN     --------+--------+--------+--------+--------+
          0 I  538  I   39  I   77  I    6  I   660
NO          I 81.5% I  5.9% I 11.7% I   .9% I 76.7%
           'I 78.1% I 83.0% I 65.3% I100.0% I
            I    .4 I    .5 I  -1.4 I    .7 I
            +--------+--------+--------+--------+
          1 I  151  I    8  I   41  I    0  I   200
YES         I 75.5% I  4.0% I 20.5% I   .0% I 23.3%
            I 21.9% I 17.0% I 34.7% I   .0% I
            I   -.7 I   -.9 I   2.6 I  -1.2 I
            +--------+--------+--------+--------+
            Column   689      47      118       6      860
            Total   80.1%    5.5%   13.7%     .7%   100.0%
```

Chi-Square	Value	DF	Significance
Pearson	12.26351	3	.00653
Likelihood Ratio	12.92373	3	.00480
Mantel-Haenszel	4.78693	1	.02868

Minimum Expected Frequency -   1.395
Cells with Expected Frequency < 5 -   2 OF   8 (25.0%)

Statistic	Value	ASE1	T-value	Approximate Significance
Phi	.11941			.00653 *1
Cramer's V	.11941			.00653 *1
Contingency Coefficient	.11857			.00653 *1
Lambda :				
symmetric	.00000	.00000		
with WRITTEN dependent	.00000	.00000		
with PASSCHAR dependent	.00000	.00000		
Goodman & Kruskal Tau :				
with WRITTEN dependent	.01426	.00818		.00658 *2
with PASSCHAR dependent	.00630	.00486		.00102 *2

*1 Pearson chi-square probability
*2 Based on chi-square approximation

WRITTEN IS PASSWORD WRITTEN DOWN? by CHOICE HOW PASSWORD WAS CHOSEN

```
                  CHOICE                                Page 1 of 1
         Count  I
         Row Pct IMEANINGF COMB. ME PRONOUNC RANDOM C OTHER
         Col Pct IUL      ANINGFUL ABLE     OMBO.
         Std Res I     1 I      2 I     3 I     4 I     5 I  Row
                                                             Total
WRITTEN  --------+--------+--------+--------+--------+--------+
            0 I   441 I    84 I    33 I    11 I    92 I   661
NO            I 66.7% I 12.7% I  5.0% I  1.7% I 13.9% I 76.8%
              I 78.6% I 73.7% I 70.2% I 57.9% I 76.7% I
              I   .5 I   -.4 I   -.5 I   -.9 I    .0 I
            +--------+--------+--------+--------+--------+
            1 I   120 I    30 I    14 I     8 I    28 I   200
YES           I 60.0% I 15.0% I  7.0% I  4.0% I 14.0% I 23.2%
              I 21.4% I 26.3% I 29.8% I 42.1% I 23.3% I
              I   -.9 I    .7 I    .9 I   1.7 I    .0 I
            +--------+--------+--------+--------+--------+
         Column   561      114     47       19      120     861
         Total   65.2%   13.2%    5.5%     2.2%    13.9%  100.0%
```

Chi-Square	Value	DF	Significance
Pearson	6.60318	4	.15840
Likelihood Ratio	6.05353	4	.19519
Mantel-Haenszel	1.40545	1	.23583

Minimum Expected Frequency - 4.413
Cells with Expected Frequency < 5 - 1 OF 10 (10.0%)

Statistic	Value	ASE1	T-value	Approximate Significance
Phi	.08757			.15840 *1
Cramer's V	.08757			.15840 *1
Contingency Coefficient	.08724			.15840 *1
Lambda :				
symmetric	.00000	.00000		
with WRITTEN dependent	.00000	.00000		
with CHOICE dependent	.00000	.00000		
Goodman & Kruskal Tau :				
with WRITTEN dependent	.00767	.00663		.15887 *2
with CHOICE dependent	.00200	.00216		.14325 *2

*1 Pearson chi-square probability
*2 Based on chi-square approximation

- - - - - Mann-Whitney U - Wilcoxon Rank Sum W Test

CHANGES   HOW OFTEN P.W. IS CHANGED
by DIFFREM   DIFFICULTY REMEMBERING

Mean Rank   Cases

    421.18   776 DIFFREM = 0 NO
    512.42   83 DIFFREM = 1 YES
               ---
               859 Total

                          Corrected for ties
       U   W   Z   2-Tailed P
    25363.5   42530.5   -4.5301   .0000

- - - - - Mann-Whitney U - Wilcoxon Rank Sum W Test

CHANGES   HOW OFTEN P.W. IS CHANGED
by WRITTEN   IS PASSWORD WRITTEN DOWN?

Mean Rank   Cases

    430.54   660 WRITTEN = 0 NO
    430.36   200 WRITTEN = 1 YES
               ---
               860 Total

                          Corrected for ties
       U   W   Z   2-Tailed P
    65972.5   86072.5   -.0127   .9899

- - - - - Mann-Whitney U - Wilcoxon Rank Sum W Test

LOGTIMES   HOW OFTEN RESP. LOGS ON
by GUESSED   WAS PASSWORD GUESSED?

Mean Rank   Cases

    414.66   801 GUESSED = 0 NO
    524.34   37 GUESSED = 1 YES
               ---
               838 Total

                          Corrected for ties
       U   W   Z   2-Tailed P
    10939.5   19400.5   -2.7265   .0064

DIFFREM DIFFICULTY REMEMBERING by WRITTEN IS PASSWORD WRITTEN DOWN?

```
                    WRITTEN       Page 1 of 1
            Count  I
            Row Pct INO      YES
            Col Pct I                .      Row
            Std Res I     0  I     1  I Total
DIFFREM     --------+--------+--------+
           0 I    619 I    158 I   777
NO           I  79.7% I  20.3% I  90.3%
             I  93.8% I  79.0% I
             I    .9  I  -1.7  I
             +--------+--------+
           1 I     41 I     42 I    83
YES          I  49.4% I  50.6% I   9.7%
             I   6.2% I  21.0% I
             I  -2.8  I   5.2  I
             +--------+--------+
            Column    660      200      860
            Total   76.7%    23.3%   100.0%
```

Chi-Square	Value	DF	Significance
Pearson	38.49327	1	.00000
Continuity Correction	36.81606	1	.00000
Likelihood Ratio	35.00983	1	.00000
Mantel-Haenszel	38.44853	1	.00000

Minimum Expected Frequency - 19.302

Statistic	Value	ASE1	T-value	Approximate Significance
Phi	.21156			.00000 *1
Cramer's V	.21156			.00000 *1
Contingency Coefficient	.20698			.00000 *1
Lambda :				
symmetric	.00353	.03213	.10977	
with DIFFREM dependent	.00000	.00000		
with WRITTEN dependent	.00500	.04544	.10977	
Goodman & Kruskal Tau :				
with DIFFREM dependent	.04476	.01703		.00000 *2
with WRITTEN dependent	.04476	.01665		.00000 *2

*1 Pearson chi-square probability
*2 Based on chi-square approximation

- - - - - Mann-Whitney U - Wilcoxon Rank Sum W Test

LOGTIMES HOW OFTEN RESP. LOGS ON
by WRITTEN   IS PASSWORD WRITTEN DOWN?

Mean Rank   Cases

446.91   650 WRITTEN = 0 NO
350.93   198 WRITTEN = 1 YES
            ---
            848 Total

                          Corrected for ties
      U   W   Z   2-Tailed P
  49783.0 .   69484.0   -4.8843   .0000

- - - - - Mann-Whitney U - Wilcoxon Rank Sum W Test

LOGTIMES HOW OFTEN RESP. LOGS ON
by DIFFREM   DIFFICULTY REMEMBERING

Mean Rank   Cases

430.22   766 DIFFREM = 0 NO
365.19   81 DIFFREM = 1 YES
            ---
            847 Total

                          Corrected for ties
      U   W   Z   2-Tailed P
  26259.5   29580.5   -2.3016   .0214

- - - - - Mann-Whitney U - Wilcoxon Rank Sum W Test

CHANGES   HOW OFTEN P.W. IS CHANGED
by GUESSED   WAS PASSWORD GUESSED?

Mean Rank   Cases

414.40   812 GUESSED = 0 NO
662.75   38 GUESSED = 1 YES
            ---
            850 Total

                          Corrected for ties
      U   W   Z   2-Tailed P
   6412.5   25184.5   -8.6554   .0000

85

DIFFREM DIFFICULTY REMEMBERING by PASSCHAR PASSWORD CHARACTERISTICS

```
                   PASSCHAR                        Page 1 of 1
            Count  I
            Row Pct IALPHABET NUMERIC  ALPHANUM ASCII
            Col Pct IIC                ERIC            Row
            Std Res I    1  I    2  I    3  I    4  I Total
DIFFREM     --------+--------+--------+--------+--------+
               0 I   631  I    42  I    97  I     6  I   776
NO               I 81.3%  I  5.4%  I 12.5%  I   .8%  I 90.3%
                 I 91.7%  I 89.4%  I 82.2%  I100.0%  I
                 I   .4   I  -.1   I  -.9   I   .2   I
                 +--------+--------+--------+--------+
               1 I    57  I     5  I    21  I     0  I    83
YES              I 68.7%  I  6.0%  I 25.3%  I   .0%  I  9.7%
                 I  8.3%  I 10.6%  I 17.8%  I   .0%  I
                 I -1.2   I   .2   I  2.8   I  -.8   I
                 +--------+--------+--------+--------+
            Column    688      47     118       6      859
            Total    80.1%    5.5%   13.7%     .7%   100.0%
```

Chi-Square	Value	DF	Significance
Pearson	11.13325	3	.01103
Likelihood Ratio	10.17997	3	.01710
Mantel-Haenszel	7.85642	1	.00506

Minimum Expected Frequency - .580
Cells with Expected Frequency < 5 - 2 OF 8 (25.0%)

Statistic	Value	ASE1	T-value	Approximate Significance
Phi	.11385			.01103 *1
Cramer's V	.11385			.01103 *1
Contingency Coefficient	.11311			.01103 *1
Lambda :				
symmetric	.00000	.00000		
with DIFFREM dependent	.00000	.00000		
with PASSCHAR dependent	.00000	.00000		
Goodman & Kruskal Tau :				
with DIFFREM dependent	.01296	.00920		.01109 *2
with PASSCHAR dependent	.00842	.00646		.00008 *2

*1 Pearson chi-square probability
*2 Based on chi-square approximation

DIFFREM DIFFICULTY REMEMBERING by CHOICE HOW PASSWORD WAS CHOSEN

```
                  CHOICE                                   Page 1 of 1
          Count  I
          Row Pct IMEANINGF COMB. ME PRONOUNC RANDOM C OTHER
          Col Pct IUL       ANINGFUL ABLE     OMBO.             Row
          Std Res I    1 I    2 I    3 I    4 I    5 I Total
DIFFREM   ---------+--------+--------+--------+--------+--------+
           0 I   515 I   99 I   39 I   14 I  112 I   777
  NO         I 66.0% I 12.7% I  5.0% I  1.8% I 14.4% I 90.3%
             I 91.6% I 86.8% I 83.0% I 73.7% I 93.3% I
             I   .3 I  -.4 I  -.5 I  -.8 I   .3 I
           +--------+--------+--------+--------+--------+
           1 I    47 I   15 I    8 I    5 I    8 I    85
  YES        I 56.6% I 18.1% I  9.6% I  6.0% I  9.6% I  9.7%
             I  8.4% I 13.2% I 17.0% I 26.3% I  6.7% I
             I -1.0 I  1.2 I  1.6 I  2.3 I -1.1 I
           +--------+--------+--------+--------+--------+
          Column   560     114      47      19     120    860
          Total   65.1%   13.3%    5.5%    2.2%   14.0%  100.0%
```

Chi-Square	Value	DF	Significance
Pearson	12.82940	4	.01214
Likelihood Ratio	10.64185	4	..5090
Mantel-Haenszel	.24380	1	.62147

Minimum Expected Frequency -   1.834
Cells with Expected Frequency < 5 -   2 OF   10 (20.0%)

Statistic	Value	ASE1	T-value	Approximate Significance
Phi	.12214			.01214 *1
Cramer's V	.12214			.01214 *1
Contingency Coefficient	.12124			.01214 *1
Lambda :				
symmetric	.00000	.00000		
with DIFFREM dependent	.00000	.00000		
with CHOICE dependent	.00000	.00000		
Goodman & Kruskal Tau :				
with DIFFREM dependent	.01492	.01070		.01222 *2
with CHOICE dependent	.00291	.00244		.04046 *2

*1 Pearson chi-square probability
*2 Based on chi-square approximation

87

PASSSAME USE THE SAME PASSWORD by DIFFREM DIFFICULTY REMEMBERING

```
                 DIFFREM      Page 1 of 1

        Count  I
        Row Pct INO       YES
        Col Pct I                  Row
        Std Res I     0  I    1  I Total
PASSSAME        ---------+--------+
          0  I   522  I   62  I  584
             I 89.4%  I 10.6%  I 84.9%
HO           I 84.3%  I 89.9%  I
             I  -.1   I   .4   I
                ---------+--------+
          1  I    97  I    7  I  106
             I 93.3%  I  6.7%  I 15.1%
YES          I 15.7%  I 10.1%  I
             I   .4   I  -1.1  I
                ---------+--------+
        Column      619       69      688
        Total      90.0%    10.0%   100.0%
```

Chi-Square	Value	DF	Significance
Pearson	1.47716	1	.22422
Continuity Correction	1.07791	1	.29917
Likelihood Ratio	1.62098	1	.20296
Mantel-Haenszel	1.47501	1	.22456

Minimum Expected Frequency -   10.430

Statistic	Value	ASE1	T-value	Approximate Significance
Phi	.04634			.22422 *1
Cramer's V	.04634			.22422 *1
Contingency Coefficient	.04629			.22422 *1
Lambda :				
symmetric	.00000	.00000		
with PASSSAME dependent	.00000	.00000		
with DIFFREM dependent	.00000	.00000		
Goodman & Kruskal Tau :				
with PASSSAME dependent	.00215	.00303		.22456 *2
with DIFFREM dependent	.00215	.00303		.22456 *2

*1 Pearson chi-square probability
*2 Based on chi-square approximation

GUESSED WAS PASSWORD GUESSED? by WRITTEN IS PASSWORD WRITTEN DOWN?

```
                        WRITTEN      Page 1 of 1
                Count  I
                Row Pct INO      YES
                Col Pct I
                Std Res I     0  I     1  I Total
GUESSED         --------+--------+--------+
                0  I    621  I    191  I    812
NO                 I  76.5%  I  23.5%  I  95.5%
                   I  95.2%  I  96.5%  I
                   I   -.1  I    .1  I
                   +--------+--------+
                1  I     31  I      7  I     38
YES     .          I  81.6%  I  18.4%  I   4.5%
                   I   4.8%  I   3.5%  I
                   I    .3  I   -.6  I
                   +--------+--------+
                Column    652      198      850
                Total    76.7%    23.3%   100.0%
```

Chi-Square	Value	DF	Significance
Pearson	.52866	1	.46717
Continuity Correction	.28171	1	.59558
Likelihood Ratio	.55662	1	.45563
Mantel-Haenszel	.52804	1	.46743

Minimum Expected Frequency -   8.852

Statistic	Value	ASE1	T-value	Approximate Significance
Phi	.02494			.46717 *1
Cramer's V	.02494			.46717 *1
Contingency Coefficient	.02493			.46717 *1
Lambda :				
symmetric	.00000	.00000		
with GUESSED dependent	.00000	.00000		
with WRITTEN dependent	.00000	.00000		
Goodman & Kruskal Tau :				
with GUESSED dependent	.00062	.00158		.46743 *2
with WRITTEN dependent	.00062	.00158		.46743 *2

*1 Pearson chi-square probability
*2 Based on chi-square approximation

GUESSED WAS PASSWORD GUESSED? by PASSCHAR PASSWORD CHARACTERISTICS

```
                  PASSCHAR                      Page 1 of 1
            Count  I
            Row Pct IALPHABET NUMERIC  ALPHANUM ASCII
            Col Pct IIC                ERIC              Row
            Std Res I     1  I     2 I     3 I     4  I Total
GUESSED     --------+--------+--------+--------+--------+
                0  I   656  I    46 I   106 I     4 I    812
  NO             I 80.8%  I  5.7% I 13.1% I   .5% I 95.5%
                 I 96.3%  I 97.9% I 91.4% I 66.7% I
                 I   .2  I    .2 I  -.5 I  -.7 I
                 +--------+--------+--------+--------+
                1  I    25  I     1 I    10 I     2 I     38
  YES            I 65.8%  I  2.6% I 26.3% I  5.3% I  4.5%
                 I  3.7%  I  2.1% I  8.6% I 33.3% I
                 I -1.0  I   -.8 I  2.1 I  3.3 I
                 +--------+--------+--------+--------+
            Column    681       47      116        6     850
            Total   80.1%     5.5%    13.6%      .7%  100.0%
```

Chi-Square	Value	DF	Significance
Pearson	18.00528	3	.00044
Likelihood Ratio	10.70256	3	.01345
Mantel-Haenszel	9.54842	1	.00200

Minimum Expected Frequency -   .268
Cells with Expected Frequency < 5 -   2 OF   8 (25.0%)

Statistic	Value	ASE1	T-value	Approximate Significance
Phi	.14554			.00044 *1
Cramer's V	.14554			.00044 *1
Contingency Coefficient	.14403			.00044 *1

Lambda :
symmetric	.00000	.00000	
with GUESSED dependent	.00000	.00000	
with PASSCHAR dependent	.00000	.00000	
Goodman & Kruskal Tau :
with GUESSED dependent	.02118	.01921	.00044 *2
with PASSCHAR dependent	.00549	.00522	.00291 *2

*1 Pearson chi-square probability

*2 Based on chi-square approximation

GUESSED HAS PASSWORD GUESSED? by CHOICE HOW PASSWORD WAS CHOSEN

```
              CHOICE                                    Page 1 of 1
         Count I
         Row Pct IMEANINGF COMB. ME PRONOUNC RANDOM C OTHER
         Col Pct IUL       ANINGFUL ABLE     OXBO.            Row
         St/ nes I    1 I    2 I    3 I    4 I    5 I Total
GUESSED  --------+--------+--------+--------+--------+--------+
            0 I   536 I  106 I   44 I   16 I  110 I  812
  NO            I 66.0% I 13.1% I  5.4% I  2.0% I 13.5% I 95.5%
              I 96.4% I 96.6% I 93.6% I 84.2% I 94.8% I
              I   .2 I  -.1 I  -.1 I  -.5 I  -.1 I
         +--------+--------+--------+--------+--------+
            1 I    20 I    6 I    3 I    3 I    6 I   38
  YES           I 52.6% I 15.8% I  7.9% I  7.9% I 15.8% I  4.5%
              I  3.6% I  5.4% I  6.4% I 15.8% I  5.2% I
              I -1.0 I   .4 I   .6 I  2.3 I   .4 I
         +--------+--------+--------+--------+--------+
       Column   556     112      47      19     116     850
       Total   65.4%   13.2%    5.5%    2.2%   13.6%  100.0%
```

Chi-Square	Value	DF	Significance
Pearson	7.43545	4	.11459
Likelihood Ratio	5.27657	4	.26008
Mantel-Haenszel	2.14264	1	.14325

Minimum Expected Frequency - .849
Cells with Expected Frequency < 5 - 2 OF 10 (20.0%)

Statistic	Value	ASE1	T-value	Approximate Significance
Phi	.09353			.11459 *1
Cramer's V	.09353			.11459 *1
Contingency Coefficient	.09312			.11459 *1
Lambda :				
symmetric	.00000	.00000		
with GUESSED dependent	.00000	.00000		
with CHOICE dependent	.00000	.00000		
Goodman & Kruskal Tau :				
with GUESSED dependent	.00875	.01017		.11498 *2
with CHOICE dependent	.00187	.00219		.17533 *2

*1 Pearson chi-square probability
*2 Based on chi-square approximation

GUESSED WAS PASSWORD GUESSED? by WORKLOC WHERE DO YOU WORK?

```
                    WORKLOC                           Page 1 of 1
              Count  I
              Row Pct IPRIVATE  HOME      PUBLIC T OTHER
              Col Pct INPS OFFI            ERMINAL          Row
              Std Res I     1 I     2 I     3 I     4 I Total
GUESSED       --------+--------+--------+--------+--------+
          0   I    131 I    177 I    487 I     15 I   810
 NO           I 16.2% I 21.9% I 60.1% I  1.9% I 95.5%
              I 86.2% I 96.7% I 98.0% I 93.8% I
              I  -1.2 I    .2 I    .6 I   -.1 I
              +--------+--------+--------+--------+
          1   I     21 I      6 I     10 I      1 I    38
 YES          I 55.3% I 15.8% I 26.3% I  2.6% I  4.5%
              I 15.8% I  3.3% I  2.0% I  6.3% I
              I   5.4 I   -.8 I  -2.6 I    .3 I
              +--------+--------+--------+--------+
      Column      152      183      497       16      848
      Total      17.9%    21.6%    58.6%     1.9%   100.0%
```

Chi-Square	Value	DF	Significance
Pearson	38.75672	3	.00000
Likelihood Ratio	29.97225	3	.00000
Mantel Haenszel	28.68754	1	.00000

Minimum Expected Frequency -   .717
Cells with Expected Frequency < 5 -   1 OF   8 (12.5%)

Statistic	Value	ASE1	T-value	Approximate Significance
Phi	.21378			.00000 *1
Cramer's V	.21378			.00000 *1
Contingency Coefficient	.20906			.00000 *1
Lambda :				
symmetric	.02828	.01397	1.98022	
with GUESSED dependent	.00000	.00000		
with WORKLOC dependent	.03134	.01561	1.98022	
Goodman & Kruskal Tau :				
with GUESSED dependent	.04570	.01923		.00000 *2
with WORKLOC dependent	.02008	.00778		.00000 *2

*1 Pearson chi-square probability
*2 Based on chi-square approximation

- - - - Kruskal-Wallis 1-Way Anova

DATAVITL   DATA IMPORTANCE
by CHOICE   HOW PASSWORD WAS CHOSEN

Mean Rank   Cases

411.94	554	CHOICE = 1	MEANINGFUL
454.83	112	CHOICE = 2	COMB. MEANINGFUL
417.15	66	CHOICE = 3	PRONOUNCABLE
599.66	19	CHOICE = 4	RANDOM COMBO.
429.38	117	CHOICE = 5	OTHER

---

848   Total

			Corrected for ties	
Cases	Chi-Square	Significance	Chi-Square	Significance
848	12.9778	.0114	13.9464	.0075

- - - - Kruskal-Wallis 1-Way Anova

PASSCHAR   PASSWORD CHARACTERISTICS
by DATASENS   DATA SENSITIVITY

Mean Rank   Cases

419.07	697	DATASENS = 1	NONSENSITIVE
441.80	94	DATASENS = 2	
472.78	43	DATASENS = 3	MODERATELY SENSITIVE
440.00	9	DATASENS = 4	
486.00	6	DATASENS = 5	VERY SENSITIVE

---

849   Total

			Corrected for ties	
Cases	Chi-Square	Significance	Chi-Square	Significance
849	2.8861	.5771	5.9977	.1993

- - - - Kruskal-Wallis 1-Way Anova

WORKLOC   WHERE DO YOU WORK?
by DATASENS   DATA SENSITIVITY

Mean Rank   Cases

441.02	699	DATASENS = 1	NONSENSITIVE
386.86	94	DATASENS = 2	
386.87	43	DATASENS = 3	MODERATELY SENSITIVE
154.06	9	DATASENS = 4	
102.79	7	DATASENS = 5	VERY SENSITIVE

---

852   Total

			Corrected for ties	
Cases	Chi-Square	Significance	Chi-Square	Significance
852	29.1286	.0000	37.0364	.0000

93

- - - - - Kruskal-Wallis 1-Way Anova

DATASENS   DATA SENSITIVITY
by CHOICE   HOW PASSWORD WAS CHOSEN

Mean Rank   Cases

416.87	556	CHOICE = 1	MEANINGFUL
459.75	112	CHOICE = 2	COMB. MEANINGFUL
400.95	47	CHOICE = 3	PRONOUNCABLE
537.13	19	CHOICE = 4	RANDOM COMBO.
425.47	116	CHOICE = 5	OTHER

                 ---

         850   Total

			Corrected for ties	
Cases	Chi-Square	Significance	Chi-Square	Significance
850	7.2643	.1226	16.2457	.0027

- - - - - Kruskal-Wallis 1-Way Anova

PASSCHAR   PASSWORD CHARACTERISTICS
by DATAVITL   DATA IMPORTANCE

Mean Rank   Cases

406.99	306	DATAVITL = 1	NONVITAL
397.26	138	DATAVITL = 2	
431.30	205	DATAVITL = 3	MODERATELY VITAL
470.19	71	DATAVITL = 4	
456.44	127	DATAVITL = 5	VERY VITAL

                 ---

         847   Total

			Corrected for ties	
Cases	Chi-Square	Significance	Chi-Square	Significance
847	8.0731	.0889	16.6678	.0022

- - - - - Kruskal-Wallis 1-Way Anova

WORKLOC   WHERE DO YOU WORK?
by DATAVITL   DATA IMPORTANCE

Mean Rank   Cases

497.05	306	DATAVITL = 1	NONVITAL
430.33	140	DATAVITL = 2	
438.61	204	DATAVITL = 3	MODERATELY VITAL
370.48	71	DATAVITL = 4	
255.47	128	DATAVITL = 5	VERY VITAL

                 ---

         849   Total

			Corrected for ties	
Cases	Chi-Square	Significance	Chi-Square	Significance
849	91.7949	.0000	116.6786	.0000

94

* * * A N A L Y S I S   O F   V A R I A N C E * * *

PASSNUM NUMBER OF CHARACTERS
by   DATASENS DATA SENSITIVITY

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	12.973	4	3.243	1.388	.236
DATASENS	12.973	4	3.243	1.388	.236
Explained	12.973	4	3.243	1.388	.236
Residual	1971.524	844	2.336		
Total	1984.497	848	2.340		

997 cases were processed.

* * * A N A L Y S I S   O F   V A R I A N C E * * *

PASSNUM NUMBER OF CHARACTERS
by   DATAVITL DATA IMPORTANCE

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	4.048	4	1.012	.430	.787
DATAVITL	4.048	4	1.012	.430	.787
Explained	4.048	4	1.012	.430	.787
Residual	1980.245	842	2.352		
Total	1984.293	846	2.345		

997 cases were processed.
150 cases (15.0 pct) were missing.

- - - - - Mann-Whitney U - Wilcoxon Rank Sum W Test

DATASENS   DATA SENSITIVITY
by GUESSED   WAS PASSWORD GUESSED?

Mean Rank   Cases

    418.26   810 GUESSED = 0 NO
    549.64   37 GUESSED = 1 YES

             . 847 Total

                          Corrected for ties
     U   W   Z   2-Tailed P
  10336.5   20336.5   -4.7827   .0000

- - - - - Mann-Whitney U - Wilcoxon Rank Sum W Test

DATASENS   DATA SENSITIVITY
by WRITTEN   IS PASSWORD WRITTEN DOWN?

Mean Rank   Cases

    425.92   652 WRITTEN = 0 NO
    424.11   198 WRITTEN = 1 YES
               ---
               850 Total

                          Corrected for ties
     U   W   Z   2-Tailed P
  64272.0   83973.0   -.1364   .8915

- - - - - Mann-Whitney U - Wilcoxon Rank Sum W Test

DATAVITL   DATA IMPORTANCE
by WRITTEN   IS PASSWORD WRITTEN DOWN?

Mean Rank   Cases

    438.27   651 WRITTEN = 0 NO
    378.99   197 WRITTEN = 1 YES
               ---
               848 Total

                          Corrected for ties
     U   W   Z   2-Tailed P
  55157.5   74660.5   -3.0856   .0020

-------------- S P E A R M A N   C O R R E L A T I O N   C O E F F I C I E N T S --------------

DATASENS   .1544
          N(850)
          SIG .000

          CHANGES

" . " IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

-------------- S P E A R M A N   C O R R E L A T I O N   C O E F F I C I E N T S --------------

DATAVITL   .1916
          N(848)
          SIG .000

          CHANGES

" . " IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

LIST OF REFERENCES

Ahituv, N., Lapid, Y. and Neumann, S. (1987), "Verifying the Authentication of an Information System User," Computers and Security, Vol. 6, No. 2, pp. 152-157.

Avarne, S. (1988), "How to Find Out a Password," Data Processing & Communication Security, Vol. 12, No. 2, (Spring, 1988), pp.16-17.

Barton, B. F. and Barton, M. S. (1984), "User-Friendly Password Methods for Computer-Mediated Information Systems," Computers and Security, Vol. 3, No. 3, (1988), pp 186-195.

Blalock, H. M. Social Statistics, Revised ed., McGraw Hill Inc., New York, NY, 1979.

Cooper, J. A. (1989), Computer an Communications Security, Strategies for the 1990's, McGraw Hill Inc., New York, NY, 1989.

Durr, M and Gibbs, M. (1989), "Peeling Back the Layers," Byte June 1989, pp. 258-259.

Fisher, R. P., (1984), Information Systems Security, Prentice- Hall Inc., New York, NY., 1984.

Hoffer, J. A. and Straub, D. W., (1989), "The 9 to 5 Underground: Are you Policing Computer Crimes," Sloan Management Review, Summer 1989, pp. 35-43.

Hoffman, L. J., Modern Methods for Computer Security and Privacy, Prentice-Hall Press, New York, NY., 1977.

Hsiao, D. K., Computer Security, Academic Press, New York, NY., 1979.

Kochanski, M. (1989), "How safe is it?," Byte, June 1989, pp. 257-264.

Kurzban, S., (1983), "A Dozen Gross 'MythConceptions' about Information Processing," Security, IFIP, pp. 15-25, 1983.

Mansfield, E., Statistics for Business and Economics, 2nd Ed., W. W. Norton and Company Inc., New York. NY, 1983.

INITIAL DISTRIBUTION LIST

	No. Copies
1. Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2. Library, Code 0142 Naval Postgraduate School Monterey, California 93943-5002	2
3. Department Chairman, Code AS Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943-5000	1
4. Prof. William J. Haga, Code AS/Hg Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943-5000	1
5. Prof. Moshe Zviran, Code AS/Zv Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943-5000	1
6. Computer Technology Curricular Office, Code 37 Naval Postgraduate School Monterey, California 93943-5000	1
7. Director Computer Center, Code 0141 Naval Postgraduate School Monterey, California 93943-5000	1
8. LT Darren A. Sawyer, U.S. Navy 3621 Selkirk Drive Winston-Salem, NC, 27105	1

Morris, R. and Thompson, K. (1979), "Password Security: A Case History," Communications of the ACM, Vol. 22, No. 11, pp. 594-597.

Menkus, B. (1980), "Understanding the Use of Passwords," Computers and Security, Vol 7, No. 2, (April 1988), pp. 132-136.

Norusis, M. J., SPSS-X Introductory Statistics Guide for Release 3, SPSS Inc., Chicago, Il., 1988.

Panns, R. and Herschberg, I. S. (1987), "Computer Security: The Long Road Ahead," Computers and Security, Vol. 6, No. 5, (1987), PP. 403-416.

Parker, D.B. and Nycum, S. H. (1984), "Computer Crime", Communications of the ACM, Vol. 22, No. 4, (1984), pp. 313-315.

Porter, J. H. and Hamm, R. J., Statistics: Applications for the Behavioral Sciences, Brooks/Cole Publishing Company, Monterey, CA, 1986.

Porter, S. N. (1982), "A Password Extension for Human Factors," Computers and Security, Vol. 1, No. 1, (1982), pp. 54-56.

Rash, W. (1989), "In Depth Security," Byte, June 1989, pp. 254.

Siegel, S. and Castellan, J. N., Nonparametric Statistics for the Behavioral Sciences, 2nd Ed., McGraw Hill Inc., New York, NY., 1988.

Smith, S. L. (1987), "Authenticating Users by Word Association," Computers and Security, Vol. 6, No. 6, (1987), pp. 464-470.

Ware, W. H. (1984), "Information System Security and Privacy," Communications of the ACM, Vol. 27, No. 4, (1984), pp. 315-321.

Wood, H. (1977), "The Use of Passwords for Controlled Access to Computer Resources," Institute for Computer Science and Technology to Computer Resources, Government Printing Office, Washington, D.C., 1977.

Woods, C. C. (1983), "Effective Information System Security with Password Controls," Computers and Security, Vol. 2 No. 1, (1983), pp. 5-10.